2015

Thermodynamics and structure of methionine enkephalin using the statistical temperature molecular dynamics algorithm

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

THERMODYNAMICS AND STRUCTURE OF METHIONINE ENKEPHALIN USING THE STATISTICAL TEMPERATURE MOLECULAR DYNAMICS ALGORITHM

by

SHANADEEN CRYSTAL BEGAY

B.S., Northern Arizona University, 2006

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2015

Approved by

First Reader

Thomas F. Keyes, PhD
Professor of Chemistry

Second Reader

John E. Straub, PhD
Professor of Chemistry

Third Reader

Ksenia B. Bravaya, PhD
Assistant Professor of Chemistry

It seems that one of the most important things in life
is to maintain consistency with one's own life goals.

To keep life balanced and yet moving forward
toward the reward of greatest value.

Such a wonderful thing to learn to manage the self!
Shanadeen Begay

Acknowledgments

It is my greatest pleasure to work within the research community with all members of the scientific community. Many of my colleagues, friends, family members, and arch-enemies have given me experiences and opportunities for personal growth through this time. I am thankful for the times we laughed, talked of our ambitions, gave emotional support, and cared for each other as fellow participants in this journey. Secondly, I appreciate all the lab members and department members whom are patient to explain to students, of all experience levels, how to approach their work. I, and many others, have found interactions with their Dissertation Advisory Committee members deeply rewarding towards the progression of each person's studies. As such, I would like to thank the Professors in my Field that were always encouraging of their students' development for being able to nurture students in equal training experiences and career enhancing relationships.

To my mentors near and far, I am indebted to your service that allows me to learn the applications of methodology to diverse fields within the framework of my dissertation. Thank you for offering appropriate advice and guidance. I dedicate my years of scientific service and outreach in your honor.

It has been an honor to work with my advisor, Professor Thomas Keyes. I look forward to continuing to build upon our work into the future. The longer I know Tom the more I emulate his research style; I respect him deeply. Prof. David Coker has always been a communal, integral, and vital mentor to my continual development as a scholar. Prof. John Straub has been helpful in a many of the steps throughout the PhD process and editing various aspects of this study – his professionalism, character, and kindness is greatly appreciated by all students I've known. Prof. Sandor Vajda has shown interest in this work and I look forward to future discussions about how to move this forward into the relm of greater relevance to structure vs. function studies. Assoc. Prof. Linda Doerrer has guided my comparisons to experimental conditions and helped me to consider all aspects of designing this work. She and John furthered official aspects of this writing. Assist.

Prof. Ksenia Bravaya and Prof. Sean Elliott are wonderful additions to the critical readers and supportive committee members. I look forward to our future endeavors. Profs. Feng Wang, Yu Xia, Alvaro Monteiro, David Wales, Peter Ortoleva, Irwin Oppenheim, other Profs. that mentored me prior to the PhD journey, Robert Best, post-docs at BU, fellow graduate students, along with a global community of researchers have helped to shape early outcomes to my results. Memories of great importance and too numerous to list here are fond milestones in my journey towards academic progress.

THERMODYNAMICS AND STRUCTURE OF METHIONINE ENKEPHALIN USING THE STATISTICAL TEMPERATURE MOLECULAR DYNAMICS ALGORITHM

(Order No.)

SHANADEEN CRYSTAL BEGAY

Boston University Graduate School of Arts and Sciences, 2015

Major Professor: Thomas F. Keyes, Professor of Chemistry

ABSTRACT

Kim, Straub, and Keyes introduced the statistical temperature molecular dynamics (STMD) algorithm to overcome broken ergodicity by sampling a non-Boltzmann flat energy histogram as noted in Kim, Straub, and Keyes, Phys. Rev. Lett. 97: 050601 (2007). Canonical averages are calculated via reweighting to the desired temperature. While STMD is promising, its application has been almost entirely to simple or model systems. In this dissertation the implementation of STMD into the biosimulation package CHARMM is used to simulate the methionine enkephalin pentamer peptide with a methione terminal cap in a droplet of CHARMM TIP3P water molecules.

Chain thermodynamics is analyzed from the novel perspective of the statistical temperature as a function of potential energy, $T_S(U)$, automatically generated by STMD. Both the minimum in the slope of $T_S(U)$, and the peak in the heat capacity as a function of temperature, calculated via reweighting, indicate a collapse transition at $T_\theta \approx 253$K. Distributions of dihedral angles are obtained as a function of temperature. Rotamer regions found in the literature are reproduced, along with unique regions not found previously, including with advanced algorithms, indicating the power of STMD enhanced sampling.

Contents

List of Tables

List of Figures

List of Abbreviations

AB	.....	A sequence label; the ith residue along a protein backbone labled as A or B.
AMBER	.....	Assisted Model Building with Energy Refinement: a version from 1995
BLN	.....	hydrophobic (B), hydrophilic (L), neutral (N)
CHARMM	.....	Chemistry at HARvard Molecular Mechanics
DOS	.....	Density of States
DR	.....	Disordered Regions
Gly-2	.....	Glycine as the second amino acid of a sequence
GREM	.....	Generalized Replica Exchange Method
LJ	.....	Lennard-Jones
MC	.....	Monte Carlo
MD	.....	Molecular Dynamics
MDLD	.....	Molecular Dynamics Langevin dynamics
Met-5	.....	Methionine as the fifth residue of a sequence
Met-Enk	.....	Methionine Enkephalin (Tyr-1, Gly-2, Gly-3, Phe-4, Met-5)
MMTSB	.....	Multiscale Modeling Tools for Structural Biology
MRR	.....	Multiple Replica Repulsion
MUCA	.....	Multicanonical Algorithm
MUCAREM	.....	Multicanonical Replica Exchange Method
NPT	.....	The isothermal-isobaric ensemble with the total number of particles (N), the constant pressure (P), the constant temperature in the system (T).

NVE	.....	The microcanonical ensemble with the total number of particles (N), the system's volume (V), the total energy in the system (E) as constants.
NVT	.....	The canonical ensemble with the total number of particles (N), the system's volume (V), the absolute temperature (T).
OP	.....	Order Parameter
PCA	.....	Principle Component Analysis
PDF	.....	Probability Distribution Function
PEL	.....	Potential Energy Landscape
pH	.....	A measure of the acidity or basicity of an aqueous solution.
Phe-4	.....	Phenyl alanine at the fourth position of the Met-Enk sequence
RE	.....	Replica Exchange
REM	.....	Replica Exchange Method
REMUCA	.....	Replica Exchange Multicanonical
STMD	.....	Statistical Temperature Molecular Dynamics
TIP3P	.....	CHARMM's Transferrable Intermolecular Potential Three-point
Tyr-1	.....	Tyrosine at the first position of the peptide backbone structure
U	.....	Potential energy
VV	.....	Velocity Verlet

WL	.....	Wang-Landau
WHAM	.....	Weighted Histogram Analysis Method
YGGFM	.....	Met-Enk; Tyrosine is"Y", Glycine is "G", Phenylalanine is "F", Methionine is "M"

Chapter 1

Computer Simulation of Complex Systems

1.1 Conventional Simulation

1.1.1 Basic Principles in STMD-CHARMM

Computer simulation is now an essential method of physical chemistry. The system is described by a potential energy function and, for bulk materials, periodic boundary conditions. The primary techniques are Monte Carlo (MC) [1] and molecular dynamics (MD) [2].

Standard MD and MC techniques

Standard MC generates a chain of configurations sampled from the canonical (NVT) ensemble, with sampling weight $W_{cano}(U,T) = e^{-\beta U}$, where U is the potential energy, $\beta = 1/k_B T$ and k_B is Boltzmann's constant. Trial moves are made by random displacements of each atom, and a move is accepted with probability $\min[1, e^{-\beta \Delta U}]$, where ΔU is the energy change due to the trial move. If the move is rejected, the old state is counted again in the chain.

For the isothermal-isobaric (NPT) ensemble [3], moves that change the volume are also attempted, and the acceptance probability is $min[1, e^{-\beta \Delta H + (N+1) ln(V'/V)}]$, for both types of move, where $H = U + PV$ is the enthalpy. Grand canonical MC, with varying numbers of particles, is also well known.

Molecular dynamics for the microcanonical (NVE) ensemble simply integrates Newton's equations, numerically, to obtain the chain of configurations. Several thermostats and barostats have been devised [4–6] allowing MD simulation of the (NVT) and (NPT) ensembles.

Equilibrium averages in the indicated ensemble of a function of the coordinates, A, are taken over the chain of configurations, or trajectory, as

$$\langle A \rangle = \frac{1}{M} \sum_{i=1}^{M} A_i. \tag{1.1}$$

where A_i is the value of the variable in the $i'th$ configuration and M is the number of configurations in the ensemble. MC is limited to time-independent averages, while MD can also describe dynamical properties.

1.2 Failures of Conventional Methods

Traditional MC and MD work well if the system is small and simple, but can fail otherwise. The significant dynamic energy range increases with system size, and the simulation must traverse the full range many times for good sampling, so, if any system is large enough, it will exceed the capability of current hardware and software infrastructure.

The state-of-the-art in supercomputing is a specialized machine, ANTON, that has simulated a millisecond of a MD trajectory. Although this is 100 times longer than any previous simulation, there are still biological systems larger than proteins that cannot yet be simulated due to power limitations and algorithmic difficulties [7]. The exascale machines, which should be able to handle biologically relevant super-systems with folding, are expected by the year 2020.

Conventional MD cannot measure long-time dynamics in explicit water without time-consuming evaluations of atomic forces. The explicit solvent MD computational time t scales, at best, linearly with the number of atoms N, at worst $N^2 \sim t$. Cutoffs curb the issue but solvent viscosity, friction coefficients, and microscopic densities are approximated by these absent and unmeasured nonbonded interactions.

1.2.1 Complex Energy Landscapes

Smaller systems are challenging if they have kinetic traps, complex energy landscapes with multiple free energy minima connected by high energy barriers [8, 9], or are in the vicinity of a phase transitions. MD will take too long to cross the barrier as will MC, due to too many rejected trial moves. Similar considerations hold for low temperatures. Organization

into a configuration with few microstates is inherently slow as so many degrees of freedom must attain special values, i.e an entropic barrier, as in protein folding [10, 11]. The result is broken or quasi-ergodicity, with the system trapped in one part of the significant configuration space, and time averages and ensemble averages are not equivalent.

1.2.2 Sampling Disordered Domains

Some protein regions are intrinsically disordered; they lack unique structure under physiological conditions [12]. The first physiologic study on these domains was in 1989 [13]. In other early studies, negatively charged groups, a large net charge at neutral pH, and a low content of hydrophic amino acid residues has also been used to characterize several biologically active disordered proteins [14]. Little is known about the thermodynamics of microseconds to minute fluctuations in disordered domain activation for any chemical system. Domain alignment matching for a full-length ID N-terminal domain of human glucocorticoid receptor and two of its naturally occurring translational isoforms [15] is one of the earliest studies on disordered regions (DR).

Standard MC and MD cannot sample the thermodynamics of the binding process in disordered domains to provide a simple mechanism for inducibility by phosphorylation. Induced phosphorylation can produce helical populations from DR that fold with substantial entropic cost. The favorable hydrogen bonding made by phosphoryl groups of polar amino acids may be driven by large enthalpy changes [16]. MD simulations of the complexation of a structured binding domain require modeling of a coil-to-helix folding transitions on the millisecond to microsecond time scale [17, 18].

Furthermore, the lack of complete mechanistic studies describing the details of binding regions/sites ambiguates parallels made between DRs and globular proteins [19, 20]. Modeling the regulation of protein functions in disordered domains with unique functional modes and comparing these findings to folding as seen in Met-Enk would further benchmark necessary structure and observables. An outstanding challenge is to reproduce physiological conditions seen in laboratory experiments [21].

1.3 Enhanced Sampling Techniques

Enhanced sampling methods such as multicanonical or entropic sampling [22], Wang-Landau random walk algorithm [23], parallel tempering or replica-exchange method [24] have been attempted to overcome the quasi-ergodic problem. Other methods are: multi-canonical replica-exchange method (REMUCA) [25], multiple histogram methods [26], and simulated tempering [27]. Simulated annealing performs a free random walk in temperature space [28]. Multi-canonical algorithim (MUCA) samples a wide phase space [29] by using a non-Boltzmann weight factor. A free random walk in potential energy space is realized. The weighting factor is determined by short trial simulations, thus, complex systems fair better with iterative convergence acceleration [22]. A combination of RE and MUCA is done to enhance efficency creating REMUCA reducing the trial simulations by at least an order of magnitude [30]. REMUCA runs need a short RE simulation that produces a weight factor used in a regular MUCA simulation.

1.3.1 MUCA and WL

Two relevant established generalized ensemble methods used to simulate protein systems are the multicanonical algorithm (MUCA) and the Wang-Landau (WL) sampling, which seek to sample a flat potential energy distribution. The distribution, $P(U)$, obeys

$$P(U) \propto W(U)\Omega(U) \tag{1.2}$$

where $W(U)$ is the sampling weight, U is the potential energy, and $\Omega(U)$ is the density of states. Thus, $W \propto 1/\Omega$ gives the flat distribution.

Each potential energy value has equal probability, giving rise to a random walk in potential energy space that will cross high-energy barriers. The density of states (DOS) is not known *a priori* and must be determined in the simulation to get the weight. The procedures are non-trival and MUCA requires tedious trial simulations. Fortunately, WL sampling employes a dynamic modification scheme for the DOS which is much faster than

the recursive refinements of the MUCA.

1.3.2 Replica Exchange

The Replica Exchange method (REM) exchanges neighboring replicas with different sampling weights for better and faster sampling. The generalized ensemble exchanges between neighbors are accepted with probability

$$A(i \to j) = \min[1, e^{\Delta_{ij}}] \tag{1.3}$$

where $\Delta_{ij} = \frac{W_i(X')W_j(X)}{W_i(X)W_j(X')}$, W_j being the sampling weight of the ith replica of configuration X. In the case of canonical replicas the expression becomes

$$\Delta_{ij} = (\beta_j - \beta_i)[U(X') - U(X)] \tag{1.4}$$

The hindering computational difficulty of REM comes from the $\sqrt{f}$ increase in the number of replicas for a proportional increase in the degrees of freedom, f, to maintain sufficient energy overlaps between neighboring replicas. Sugita and Okamoto combined the REM and MUCA methods to create multicanonical replica-exchange method (MUCAREM) and replica-exchange multicanonical (REMUCA) method [25]. This complicated method to overcome the poor scaling of REM still requires much experience in determining the weight for the MUCA portion of the algorithm. An alternative to the REM method that utilizes STMD is the Generalized Replica Exchange Method (GREM) [31]. Regular parallel-tempering, another name for replica exchange, gives weights that are *a priori* known and this is an advantage of the technique [32]. Thus, the application of this method is easier than other generalized ensemble methods and is easily combined with other techniques.

For conventional MC using a replica exchange method, the global acceptance probability depends on the difference in energy between independent replicas. A low acceptance probability can result if large energy differences between replicas exist [33].

1.3.3 Hamiltonian REM

A modification to the REM method is the Hamiltonian REM [34] that is based on the modification of the Hamiltonian. The Hamiltonian forces can be scaled making the interactions of interest prominent and enforcing an effective temperature of a partial system. If a partial system's effective temperature scaling is small relative to the total degrees of freedom, the number of replicas needed is reduced. For globular proteins a combination of side chain, reconfiguration, or entanglement interactions are responsible for the ruggedness of the potential energy landscape. By weakening the part of the Hamiltonian protein interactions that create ruggedness, a flatter energy landscape is more readily explored. Through scaling part of the Hamiltonian, an efficient sampling is possible on a smoother potential energy landscape compared to the real one.

Hamiltonian REM has been extended to "solute tempering" [35] for the case of the solvated alanine dipeptide, in which the potential energy is scaled with each temperature replica instead of scaling the temperature only. High temperature replicas of an alanine dipeptide are sampled frequently. The low-temperature 512 water molecule replicas stay near the target temperature. As a result, the number of replicas is reduced due to independence from solvent-solvent interaction energy, which is the main factor that leads to poor scaling with system size in ordinary REM. The resulting acceptance probability for replica exchange scales with the number of degrees of freedom of the biomolecule and not with the number of water molecules. The STMD method can be modified to adopt this strategy to overcome the scalability problem of the conventional REM. Solvation STMD, which is based on the dynamical update of the statistical temperature as a function of the partial energy composed of (i) protein-protein, (ii) protein-water, and (iii) water-water interactions, can accelerate the convergence of simulations.

1.4 Principles of STMD

The multicanonical MD and WL sampling techniques are foundational to STMD via the one-to-one mapping between the statistical temperature, $\tilde{T}(U)$, and $\Omega(U)$ using non-

deterministic sampling.

1.4.1 Principles of Computer Simulation

The total energy, TE, is the sum of the kinetic energy, KE, and the potential energy, U,

$$TE = KE + U. \tag{1.5}$$

The potential energy distribution, $P(U)$, for a computer simulation, obeys Eq. 1.2.

The Boltzmann weight, $e^{-\beta U}$, yields a canonical ensemble; the result is a Gaussian distribution. The peak in this distribution may be too narrow to allow the system to cross barriers and achieve good sampling. Thus, generalized ensembles using alternative sampling weights are of interest.

Most experiments measure properties that are canonical averages. Reweighting techniques allow calculations of canonical averages from the generalized ensemble.

One specific choice is the WL weight $W_{wl}(U)$,

$$W_{wl}(U) = \frac{1}{\Omega(U)}, \tag{1.6}$$

which yields a flat-energy distribution which gave a flat-energy Monte Carlo (MC) algorithm including a method to calculate the unknown $\Omega(U)$ on the fly. Therefore, a flat-energy MD algorithm is desirable, STMD.

A canonical MD simulation yields the Boltzmann sampling weight. Consider a canonical simulation at temperature T_0 where the potential energy function is replaced with an effective potential

$$v(U) = T_0 S(U), \tag{1.7}$$

where $S(U)$ is the entropy. This effective potential is used in our statistical temperature molecular dynamics (STMD) algorithm. The STMD weight is then expressed as

$$W(U) = e^{-S(U)/k_B} \tag{1.8}$$

Since

$$S(U) = k_B ln\Omega, \tag{1.9}$$

$$W(U) = \frac{1}{\Omega(U)}, \tag{1.10}$$

Thus STMD, as WL, samples the flat-energy distribution.

1.4.2 Statistical Temperature Dynamic Update

The dynamic, or on-the-fly, update scheme of the sampling weight in STMD [36] is based on the basic thermodynamic relationship between the entropy $S(U)$ and the statistical temperature $T(U)$ of

$$T(U) = \left[\frac{\partial \ln \Omega(U)}{\partial U}\right]^{-1} = \left[\frac{\partial S(U)}{\partial U}\right]^{-1}. \tag{1.11}$$

Likewise, the statistical temperature estimate can be made with Eq. 1.11.

Specifically, a flat energy distribution is obtained by refining the statistical temperature estimate, $\tilde{T}(U)$, instead of the DOS estimate, $\tilde{\Omega}(U)$. The DOS is related to the microcanonical entropy as $\tilde{S}(U) = \ln \tilde{\Omega}(U)$ and $k_B = 1$. The instantaneous estimates of the statistical temperature $\tilde{T}(U)$ are defined on an energy grid. The jth position in the grid is $U_j = G(U/\Delta)\Delta$ with bin size Δ and a function $G(x)$, which returns the nearest integer to x.

The WL sampling idea modified the DOS by multiplying the existing DOS by a modification factor f (where $f > 1$) every time the system visits U_j. Inclusion into STMD yields $\tilde{\Omega}_j \to f\tilde{\Omega}_j$, which reduces to $\tilde{S}_j \to \tilde{S}_j + \ln f$. The dynamic update scheme for the statistical temperature estimate $\tilde{T}(U)$ relies on this advantage.

The entropy update combined with the central finite difference approximation of

$$\frac{\partial \tilde{S}}{\partial U}\Big|_{U=U_j} = \frac{1}{\tilde{T}(U_j)} = \frac{\tilde{S}_{j+1} - \tilde{S}_{j-1}}{2\Delta}, \tag{1.12}$$

gives the dynamic update scheme for the statistical temperature

$$\tilde{T}'_{j\pm1} = \frac{\tilde{T}_{j\pm1}}{1 \mp \delta f \tilde{T}_{j\pm1}} \tag{1.13}$$

where

$$\delta f = \frac{\ln f}{2\Delta} << 1 \;\; and \;\; f \geq 1 \tag{1.14}$$

Equation (1.13) gives a natural increase in temperature as the increasing slope of the value of $\tilde{T}_{j+1}$ and decreasing slope of $\tilde{T}_{j-1}$are found successively. Finally, the restricted temperature sampling generates a canonical ensemble with temperatures of T_L and T_H in both low and high energy ends. For further details on mathematical implementation, a paper by Kim, Straub, and Keyes [37] is useful.

The range of statistical temperatures is selected between an upper T_H and a lower T_L. Canonical simulations at T_H and T_L indicate the corresponding energy range, and energy bins are set up to extend well below and well above the lower and upper energies, respectively. While there are many possibilities, currently the kinetic temperature controlling the momenta, T_0, is equated to T_H, and $T(U)$ is initialized as the constant $\mathrm{T}(U) = T_H$.

The rule for updating $T(U)$ is, upon a visit to energy bin j with corresponding discretized statistical temperature T_j,

$$\frac{1}{\tilde{T}'_{j\pm1}} = \frac{1}{\tilde{T}_{j\pm1}} \mp \delta f \tag{1.15}$$

where δf is defined in Eq. (1.14)

and the convergence factor, f, is written

$$f = 1 + f_d \tag{1.16}$$

where $f_d << 1$ is the modification factor. Temperatures below T_L and above T_H are not allowed; this results in a canonical simulation at these low and high energy extremes.

The final probability distribution function (PDF) is obtained as

$$P(U) \approx e^{S(U)-\tilde{S}(U)} = e^{\int^U \delta\beta(U')dU'} \tag{1.17}$$

where

$$\delta\beta(U) = \frac{\tilde{T}(U) - T(U)}{(\tilde{T}(U)T(U))}. \tag{1.18}$$

If $T(U) = \tilde{T}(U)$, then the PDF is a constant and the energy distribution is perfectly flat. The entropy estimate, $\tilde{S}(U)$, is calculated by integrating the inverse of the linearly interpolated statistical temperature as

$$\tilde{S}(U) = \int_{U_l}^{U} \frac{dU'}{\tilde{T}(U')} = \sum_{j=l+1}^{i^*} \frac{\Delta}{\left(\tilde{T}_j - \tilde{T}_{j-1}\right)} \ln\left[1 + \left(\frac{\tilde{T}_j}{\tilde{T}_{j-1}} - 1\right)\frac{(U - U_{j-1})}{\Delta}\right] \tag{1.19}$$

where $U_i = (U_i + U_{i+1})/2$ as the midpoint between two grid points and $i^* = i - 1$ if $\bar{U}_{i-1} \leq U \leq U_i$ or $i^* = i$ if $U_i \leq U \leq \bar{U}_i$.

An STMD simulation is divided into stages, starting with Stage 1. The f_d is initialized to the "large" value of 10^{-4}. Periodically a dig operation is performed, in which the $T(U)$ is replaced by the lowest value attained for all energies below that value. Stage 1 ends when $T(U)$ finds T_L, then Stage 2 begins. An accumulated energy histogram is periodically checked for flatness. When a specified flatness is attained, the convergence factor is reduced by the substitution $f \rightarrow \sqrt{f}$. Stage 2 ends when a target value of f_d, 10^{-8} for this case, is reached. Finally the production run begins; Stage 3 produces the data used to calculate observables and thermodynamic properties. Canonical averages are calculated via reweighting to the desired temperature.

The lowest and highest energies are selected to span the range of significant values sampled between T_L and T_H. The histogram flatness tolerance is either 20% or 30%; there is a binsize Δ of 25. For example, input parameters and units for a simulation with the same

system setup as discussed in Chapter 3 are: characteristic temperatures are $T_0 = 300K$, $T_L = 50K$, and $T_H = 700K$ with a lowest energy at -13000 kcal/mol and highest energy set to -1000 kcal/mol.

1.5 Demonstrated Performance of STMD

STMD has been applied to a Lennard-Jones (LJ) fluid [36], off-lattice Honeycutt-Thirumalai β-barrel hydrophobic, hydrophillic, and neutral (BLN) [37], AB [38], and other liquid crystalline systems [39]. The inherent structure analysis of an off-lattice and coarse-grained BLN 46-mer and 69-mer showed STMD to overcome the ruggedness of the energy landscape outperforming calculations executed in previous studies. The STMD sampling of the 110 particle LJ fluid showed comparison estimates of a heat capacity which confirmed previous findings. The internal energy errors were less than 0.0004 for the 110-particle LJ fluid system indicating a high degree of accuracy [37]. No units are given for the internal energy of the LJ fluid.

The BLN 46-mer system STMD study [37] found a new temperature of folding, T_f=0.2, which is below the previously reported REMC results (T_f as 0.27). A difference of 0.07 shows the superior sampling of STMD; no units are given for the ST. Complex PEL are also well sampled by STMD; low-lying states are easily seen as clear funnels. The 69-mer shows an inherent structure (IS) scatterplot with new insights into the T_f via the intermediate characteristic temperature of $T_{p_0} = 0.34$. The inherent structure (IS) scatterplot of the 46-mer shows low-lying states IS_1-IS_5 in the multifunnel landscape which has not been reported previous to this study.

The indicated global minimum in the Kim, Straub and Keyes study [37] is IS_0. IS_0 is populated much less than the broad mis-folding funnels IS_1 and IS_2 for non-native IS's causing broken ergodicity in regular MD. The 69-mer has a dominant mis-folding funnel which is more populated than the funnel with the global minimum. All lowest energy basins are energetically similar indicating the high frustration and small conformational changes, upon minimization, in a folding funnel. All these IS PEL details give new insight into

the folding traps that cannot be characterized with traditional MD/MC methods. Similar studies with competing techniques [40] also found the global minimum.

The STMD algorithm has shown the uniform sampling of the potential energy space in an isotropic-liquid crystal phase transition in liquid crystalline systems. The temperature evolution of the potential energy and the orientational order parameter of a spherocylinder was in excellent agreement with previous studies of conventional const-NVT MD simulations in the Lintuvuori and Wilson paper [39].

1.6 Advantages of STMD

Compared to other generalized enhanced sampling algorithms the unique advantages are:

1. The determination of the sampling weight is fully automatic via the dynamics update scheme for the statistical temperature.

2. The use of simple force modification $\tilde{\mathbf{f}}_i = \gamma(U)\mathbf{f}_i$ allows easy implementation of STMD into ordinary MD simulations.

1.7 Combination of STMD with CHARMM

The STMD algorithm is implemented in a local verison of The Chemistry at HARvard Molecular Mechanics (CHARMM) package [41]. The files used to setup the system were CHARMM standard input files using customized changes. The STMD enhanced CHARMM is custom built in CHARMM 32 using the GNU compiler and each run was done on one IBM LS21 blade server machine with dual-core 2.6 GHz AMD Opteron 2218HE processor and the BU-Linux operating system. The version of the executable for CHARMM 36 is consistent with previous studies using STMD. These studies were meant to compare with the same force-field, version changes, and starting configuration files as Nayouki Miyashita and Jaegil Kim's studies for direct comparison at the machine level.

Chapter 2

Met-Enkephalin Computational Approaches

2.1 Computational Observables in Met-Enkephalin Studies

Met-enkephalin (Met-Enk) has become a standard "test system" for new algorithms and potential energy functions. The first MC simulation by Paine and Scheraga [42] found low-energy conformations using an empirical potential [43]. For each of the naturally occuring amino acids empirically determined data is used to define partial atomic charges, hydrogen bond energies, nonbonded interaction energies, geometric parameters, and intrinsic torsional potentials [44]. Empirical interatomic potentials [41] such as the Empirical Conformational Energy Program for Peptides (ECEPP/2) [45] were used with different molecular dynamics environments to determine a set of conformations. EECPP/2 can be used with solvent effects which must be added seperately. Small peptide benchmarking also uses different enhanced sampling algorithms to focus on gathering, organizing, and refining computational data in comparison to experimentally observed properties. All studies that report the collapse temperature T_θ and the temperature of folding T_f include implicit or vacuum environments except aqueous solution results by Kinoshita et al. [46, 47] and Mitsutake et al. [48, 49]. These studies use a reference interaction site model and various simulation methods.

Determination of order parameters (OPs) from the results of a system sampled via accelerated or enhanced sampling algorithms is standard practice. A common theme is to compare clusters of structures that inform a particular subset of molecular functionality. Unique features defining a group behavior can be related to a particular molecular setup leading to a defining Function vs. Structure principle. For example, indexing and categorizing conformational groupings and comparing changes in helicity over time inform structural OPs. Measures of compactness of backbone structures give insight into foldability [50]. Similarity measures comparing two states: one reference and one along a continuously defined variable, relate chemically relevant interactions. Native contacts are

an example of a similarity measure. These can be used to assess frustration of a system where attraction between residues out of contact in the native state (i.e. global minimum) are neglected.

Collective variables [51], such as disconnectivity graphs, measured at relevant points in the trajectory of dynamical variables, such as along the intermediate states on the transition pathway, can inform about the relative heights of barriers found in the energy landscape. Other measurable indicators such as categorizing the dihedral angles (φ) [52–54], end-to-end measurements [55, 56], and PCA analysis [53, 54] elucidate structural OP benchmarks.

Average properties help determine the interpretable results for other OPs. Some average properties specify the temperature of folding [55, 57, 58], the collapse temperature [55], and the heat capacity, $C(T)$, [59] to show relative equilibrium measures for the Met-Enk system. Numerical parameters such as the number of configurations per energy basin, the ground state inherent structure count, and a global minimal energy configuration all set boundaries on the folding landscape of Met-Enk. One way to probe the kinetic pathways in the energy landscape of Met-Enk is to use the potential of mean force [53]. Minimization and pathfinding strategies connecting minima on the energy landscape are used to find shortest transition pathways and transition rates between landscape features of relevance [59]. Another kinetic metric is using path sampling between energetically categorized configurational clusters and calculating a rate of escape from free energy basins [60].

Indicators such as the Energy vs. MD step show the efficiency of energy landscape explorations. Probabilities of properties and OPs provide metrics for comparisions between algorithmic techniques. Histograms of energies vs. time step or binning energies vs. ST indicate the distributions that realize previously discussed mathematical formulations. Temperature dependence on these variables also gives insight into thermodynamic stability and ergodicity.

A variety of methods of sampling were also tested on Met-Enk, as summerized in Table 2.1. A parallel-tempering exchange of configurations improved sampling when com-

Potential	Type	Collapse Temperature (T_θ)	Method	Folding Temperature (T_f)	Method
$ECEPP/2$ $KONF90^a$	vacuum	280±20 K	$C(T)$	230±30 K	$\frac{d\langle O\rangle T}{dT}$
$ECEPP/2$ $KONF90^a$	vacuum	310±20 K	$\frac{d\langle V\rangle T}{dT}$	none	none
$CHARMM19^{b1}$	implicit	310 K	$C(T)$	ambiguous	φ OP
$CHARMM19^{b2}$	implicit	260 K	$C(T)$	ambiguous	φ OP
$AMBER99^c$	vacuum	none	none	320K	$C(T)$
$SMMP^d$	vacuum	280K±10K	$C(T)$	none	none

Table 2.1: A vacuum study using the ECEPP/2 with KONF90$^\alpha$ potential done by Hansmann, Masuya, and Okamoto [58] in vacuum. An implict solvent study by Evans and Wales [59] has two reported T_θ and T_f temperatures: CHARMM19^{b1} and CHARMM19^{b2}. Met-Enk in vacuum results by Kim [55, 62] with potential AMBER99^c give only a folding temperature. The SMMP^d $C(T)$ collapse temperature found by Yang and Kwak [63] is in vacuum. All potentials are empirical.

pared to canonical sampling with MC updates [58]. For example, when sampling broad local minimum in the energy landscape, the parallel-tempering simulations quickly converge to the stationary distribution. A comparative MC canonical run fails to escape a local energy minimum within the larger energy basin [53]. As the temperature is changed the MC canonical run expectation values for the average energy remain unreliable due to non-thermalized simulations. The lower bound on the number of independent ground state conformers near -10.72 kcal/mol was an order of magnitude higher for parallel-tempering verses canonical MC [32]. Simulations with the MUCA approach [61] predict the tertiary structure of Met-Enk with one run of the simulation.

The unique collapse temperatures predicted are listed in Table 2.1. The original work calculating T_θ is by Hansmann, Masuya, and Okamoto; these authors found a value of 280±20K [58] from a peak in the specific heat in implicit solvent. Histogram reweighting

of Parallel Tempering Monte Carlo data has a T_θ at 310K [59] with implicit solvent. Comparisions to the free energy as a function of both peptide volume and overlap in a distance order parameter at 300K support a T_θ of 300K [57]. Hansmann et. al. have a series of papers relying on a value of 300K [57] for T_θ, performed in implicit solvent.

The folding transition is found in the temperature range of 230K±30K by Hansmann, Masuya, and Okamoto [58] which are studies in implicit and vacuum environments. T_f=230K is found from differences in the temperature that measure the derivative to the curve which measures similarity to the native state. Tsallis Statistics were used by Kim, Funkunishi, and Nakamura [55, 62] to sample the Met-Enk in vacuum with a folding transition at 320K. Yang and Kwak [63] also sample Met-Enk in implicit solvent and find a folding temperature at 380±10K.

Evans and Wales [59] find a temperature range from 200-300K using an OP defined from the dihedral angles. These OPs are calculated via the harmonic superposition approximation (HSA) and parallel tempering monte carlo (PTMC). Two estimates comprise the HSA: 1) Harmonic densities of states are used to estimate the equilibrium occupation probability of a single minimum using a superpostion approximation. 2) The harmonic densities of states also estimate the canonical partition functions. PTMC uses the minimization of the sampled instantaneous configurations at a fixed interval. If probabilities obtained from HSA exclude the global minimum and remove the folding group A, which has the lowest free energy minima, the T_f becomes 260K. If the group A and free energy minima are included the folding occurs over a range of temperature from 35K to 260K.

The Kim, Funkunishi, and Nakamura paper [55] reports an average end-to-end distance from the nitrogen of the N terminus to the oxygen of the C terminus of 6.2 Å at low energies of approximately 40 kcal/mol. The PDF obtained from the canonical ensemble using a Tsallis inspired method is identical to the canonical distribution for all regions except near the folding transition. The Tsallis inspired PDF at 400K has slight transition region deviations from canonical MD in the transition region around 350K. This PDF indicates removal of quasiergodicity with this method [55, 62].

The lowest-energy conformations were realized [61, 64, 65]. One report of the global minimum potential energy value was -10.7 kcal/mol [65] and was found from a single simulation run. The end-to-end distance of the peptide is smaller for simulations at lower energies suggesting a tight global minimum structure.

In other simulations, conformational analysis is the primary focus. The ground state conformation called Structure A has Type II β-turn between the Gly2 and Met5 residues and is prominently discussed. This turn is possibly stabilized by two hydrogen bonds between amino-acids. The second-lowest energy state in the Hansmann and Onuchic study [66] is characterized by another Type II β-turn between Tyr1 and Phe4 with a hydrogen bond. The number of configurations in a 32ns constant temperature MD simulation (N,V,T) was the same as the RE results [53]; however the RE explores five times more configurational space than MD.

The solvation free energy in SPC/E water is between 145 and 200 kcal/mol set for 298.15K in the multicanonical reference interaction site model (RISM) [48]. The total energy ranges from about 173 to 185 kcal/mol in RISM runs at 300K and from 178 to 220 kcal/mol in runs at 800K, respectively. The energies of the eight conformations vary between 175.4 and 180.3 kcal/mol. The end-to-end distance of the lowest-energy conformation is 13.6 Å and extended. At T = 100K the eight lowest-energy conformations are found to be sampled in the Gly3 and Phe4 φ. The distributions spread with increasing temperature, implying large thermal fluctuations at 800K.

2.2 Physical and Chemical Properties of Met-Enkephalin

Enkephalins aid in the immune response, pain tolerance, dependance, and inhibition [67]. In nerve tissue terminals [68] there are μ, κ, and σ opiate receptors. Enkephalins and endorphins primarily bind as ligands to sensitive σ-receptors during stress reactions in the body. The ligand-receptor binding modifies the neural signal by changing the potassium ion conduction [69]. Met-Enk originates from $\beta-$lipotropin [70] and is present in the headpiece of all larger endorphins.

Experimental studies show that Met-Enk is conformationally flexible in aqueous solutions, including the local H-bonding backbone segments [71], with ambiguous secondary structure. The only stabilized $\beta-$turn (by phenyl-phenyl hydrophobic interactions) is found in 50 mM of sodium dodecyl sulfate micelles. X-ray crystallography [72] has shown an extended conformer while both nuclear magnetic resonance (NMR) [73, 74] and molecular simulations [75, 76] find similar open conformers.

The orienting of the peptide in the membrane may restrict molecular motions [74]. When considering a mixed lipid/aqueous environment around a Met-Enk molecule, a decrease in polarity of the lipid/water environment upon transfer from an aqueous to lipid-rich environment may be observed [77]. A population of intermolecularly hydrogen-bonded conformations may be found if incremental chemical shifts can be related to specific bonded pairs of residues indicating β-turns. The conformations observed are expected to be restricted in comparison to the water experiments since the specific orientations to the membrane phase are dominant [77].

Other studies with phospholipids and Met-Enk indicate an "attraction-interaction" model of binding of emkephalin to phospholipids. The hormone is attracted by the negatively charged lipid, using electrostatic interactions. Site-specific hydrophobic contacts facilitate the energy of Met-Enk, then orient into the lipid phase. Specifically, NMR data suggests that Met-Enk binds to the lipid with Tyr-1, Phe-4, and Met-5 side-chain substituents. These rotamers are associated with nonpolar interior regions of a micelle. A COOH carboxylate on Met-Enk is on the surface of the lipid particle [74].

Work by Zetta [78] shows inter-molecular hydrogen bonds between Phe-4 and Met-5 at different pK_a values not seen in previous simulations. Nuclear magnetic resonance spectroscopy (NMR) supports the presence of a turn at Gly 3 and Phe 4 in dimethyl sulfoxide [79]. A trihydrate crystal structure [80] from a water solution reveals coiled β-III (Gly2-Gly3) and β-I (Gly3-Phy4) turns. Similarities to morphine between tyramine and cyclohexenyl rings are similar to the Tyr1 and Phe4 ring arrangement [81].

Met-Enk plays an important role in immunomodulatory signaling and regulates cy-

tokine secretion. Specifically, Met-Enk can exert positive modulation to the pathway between the DC and CD4+T cells. Met-Enk is a possible pharmacological or immunological agent to be used in vaccine preparations against life-threatening diseases like AIDS [82].

The ground state configurations of Met-Enk were all found below -10 kcal/mol. The lowest global energy found is -12.2 kcal/mol [83] simulated at 50K.

2.3 Conventional Simulations of Met-Enkephalin

A study by Hansmann and Okamoto found that to explore the global-minimum energy with high quality statistics canonical simulations require 10^{10} times more simulation time than a multicanonical algorithm [61]. Canonical MD simulations get trapped below 300K for 1 ns simulations [30]. In another simulation, an 18 ns MD simulation becomes trapped in the extended states and shallow energy wells [54]. Pure MD dihedral and torsion angle distributions rarely show sampling of all known deep wells. Few end-to-end measurements indicate compact states in explicit water at a pH < 9. As reported, a 2 ns MD simulation in explict water rapidly stabilizes into a single bend backbone conformation [84] with no futher conformational changes ending 10ns. Another all-atom molecular dynamics (MD) simulation that reports the hydrophobicity (or hydrophicility) of Met-Enk atom pairs AB described in terms of the solute−solvent and site−site radial distribution functions $g_{AB}(r)$ [85]. The 186 waters and one Met-Enk by atomic sites are classified in the following categories: highly hydrophilic, hydrophilic, hydrophobic, and exceptions arising from steric or proximity constraints. A qualitative similarity in the behavior of $g_{AB}(r)$ obtained from theory and simulations is found. Upon inspection, the structure in the first solvation peak in the solute-solvent (site-site) radial distribution funtions $g_{AB}(r)$ shows four classes of atoms. These correspond qualitatively well with the more complicated MD simulation peaks.

MC variants were designed to overcome obvious local ergodic traps seen in short MD runs. Improvements focused on hamiltonian changes, forcefield modifications optimized to explore a more complete energy landscape, and changes to partition functions to describe

a larger range of states and ensembles for a given order parameter.

2.4 Review of Sampling Algorithms of Met-Enkephalin

The variation in consistent sampling of β-turn regions of Met-Enk using different enhanced sampling techniques show remarkable variability of Gly2 and Tyr1 ramachandran angles (RA). Phe4 and Gly3 show some similarities in sampling. The orienting of the peptide in the membrane may restrict molecular motions [74]. Regular MD simulations with no solvent run at body temperature (310K) for 60 ps have some β-bend conformational changes. The $Tyr^1(CO)-Phe^4(NH)$ hydrogen bond stabilizes the β-bend conformation while $Tyr^1(NH_3^+)-Phe^4(CO)$ and $Tyr^1-(NH_3^+)-Met^5(CO_2^-)$ [86]. A population of inter-molecular hydrogen-bonded conformations may be found if incremental chemical shifts can be related to specific bonded pairs of residues indicating β-turns. The conformations observed should be restricted in comparison to the water simulations since the specific orientations to the membrane phase would be dominant. Hybrid Monte Carlo [87] and Langevin [88] are other variations of MC applied to multi-canonical ensembles tested with the Met-Enk system.

The "multiple replica repulsion" (MRR) approach [52] is similar to the replica exchange method but has a repulsive potential to minimize over-sampling of densely populated states. MRR evolves distinct replicas and biases the potential against conformational similarity between a trial and current configuration. Acceptance probabilities are described by the replusive term between replicas or by relatively small energy differences in small displacements in a single replica. Thus, the number of replicas does not scale with degrees of freedom. For MC implementations in this study, the MRR accepts trial moves frequently (near 1).

40 ns MD Langevin dynamics (MDLD) studies were done [56] of Met-Enk in explicit 186 TIP3P waters used AMBER [89]. The AMBER potential function and TINKER 3.7 molecular package [90] report a probability distribution for the radius of gyration (RG) of

Met-Enk. The square of the radius of gyration is computed as

$$R_g^2 = \frac{1}{N}\sum_{i=1}^{N}\left(r_i - r_g\right)^2 \tag{2.1}$$

where r_g is the position of the center of gravity of Met-Enk. R_g^2 is maximum at $\sim$ 60 Å^2. The extended state is within the range of 50-60 Å^2. Other dominant configurations are the semi-packed state with $R_g^2 \sim$ 20-40 Å^2 and the packed state where $R_g^2 \simeq$ 15-20 Å^2.

The Met-Enk peptide stays in a packed or semi-packed state for a majority of the MDLD simulation, rapidly interconverting. The MDLD simulation shows gauche(-/+) conformers at 29% and 34% probability respectively. The *trans* probability from MD simulation is 37% and is slightly higher than gauche(-/+) conformers. The solvent-accessible surface area potential simulations show 99.8% trans configurations [56].

Shen and Freed [56] discuss the solvation potential, the TIP3P water viscosity (62% of experimental value), and dielectric constant were measured for explict water simulations. The flipping of the Tyrosine-Tryptophan residues is measured by fluorescence anisotropy techniques and yields a P_2-type correlation function. In the solvent accessible surface area, the correlation time τ for Tyr-1 is 86 ps and 99 ps; one order of magnitude smaller than regular MD simulation results. The end-to-end vector relaxation times are 402 and 406 ps. In regular MD simulations the end-to-end relaxation time is 232 ps. The use of simple friction coefficients and a simple solvation potential allow the implicit water simulations to be two orders of magnitude faster than the explicit water simulations [56].

Su and Cukier present an explict solvent MD study of Met-Enk [54]. Two PCA modes that capture significant long-time backbone orientations in conformational space are discussed. A closed salt-bridge-like structure (at 4.5 Å) shows a stable minimum in the PMF end-to-end distance seperated by a 2 kcal/mol barrier (at 7.5Å, k_BT=0.6 kcal/mol at T=300K) from a flatter region (at 10-14 Å). The end-to-end distance of the peptide is measured using distance REM (DREM). DREM attempts to decrease the degrees of freedom that describe the system by minimizing the differences between the total number

of degrees of freedom in the potential functions. Another method studied is Hamiltonian REM (HREM). In this method, consolidated hamiltonians that describe a known reaction coordinate are combined with umbrella sampling methods to scale the peptide-peptide and peptide-solvent electrostatic and LJ potentials. Since only potential functions with a restricted set of degrees of freedom can be exchanged via RE, all solvent degrees of freedom are not included. The one dimensional PCA analysis shows two wells with a shallow barrier supporting a coexistence of conformations of Met-Enk in water with non-distinct secondary structure features. The conformations in the lowest PCA first mode well are at -0.6 to -0.4 Å [54].

The PCA diagonalizes the covariance matrix with the eigenvectors of the covariance matrix σ as $(\sigma)_{ij} =< \delta\alpha_i \delta\alpha_j >$ where the atom fluctuates as $\delta\alpha_i = \alpha_i - < \alpha_i >$. Using the Cartesian components of the position of the ith atom, $\alpha_i = x_i, y_i, z_i$, PCA decomposes the configuration point as $\mathbf{X}(t) = \sum_{i=1}^{3N} [X(t) \cdot \mathbf{m_i}]\mathbf{m_i} = \sum_{i=1}^{3N} p_i(t)\mathbf{m_i}$ where $\mathbf{m_i}$ are the orthonormal eigenvectors of the covariance matrix.

Enhanced sampling is performed along an identified reaction coordinate $\mathbf{r} = \mathbf{f}(\mathbf{X})$. Replicas that differ by a window potential are introduced at a desired distance into DREM. Using this method, there is not a large entropy decrease when Met-Enk is stretched. Analysis with DREM entropy and energy separation finds an **r**-dependant free energy $\Delta A(r) = \Delta E(r) - T\Delta S(r)$ by using weighted histogram analysis method (WHAM). WHAM combines potential energy data from histograms with different bias window potentials to estimate a probability density via a linear combination technique. A joint probability density is an extension of WHAM used for a specific value of the **r** reaction coordinate [54].

The average relative total energy and average dihedral angle conformation for four beta turn angles are calculated by Carlacci [79]. The lowest energy extended conformation of Met-Enk from X-Ray crystollography is reported as the most sampled structure [79]. The G-G Type II' angle is the lowest energy structure for the lipid-like state which includes Tyr1-B0 and B1 turns and Gly2-AL, C, AR, B0, D turns, Gly3-B, Phe4-B, Met5-X. The

solvent environment, specifically the ionization state, is considered in the aqueous MD runs with the i/i-state (zwitterionic state) and the n/n-state (lipid-like) environment. The Monte Carlo Simulated Annealing (MCSA) method has a slight modification to the trial acceptance but standard Metropolis acceptance rules apply. A trial conformation is compared to the energy of the previously accepted conformation and is accepted if the trial move is less in energy. Standard MC acceptance criterion is used in other cases. An angle perturbation restricts the random walk of any randomly chosen backbone angle. At the start of the MC procedure, a random high Boltzmann temperature factor seeds the MCSA and as the temperature drops the conformational search is restricted to lower energy pathways.

A REMD method is used to sample 16 temperature replicas from a system equlibrated at 300K in an NVT for 32 ns [53]. The approximately 600 explicit waters were equilbrated for 100 ps with periodic boundary conditions and a 9 Å cutoff in long-range force calculations. The potential of mean force is calculated by Sanbonmatsu and co-authors in [53]. Met-Enk has low energy barriers allowing for interconversion between structures. Helicity is described as being found in structure 1 while structures 2 through 4 are non-helical and interconvert. Specific details on the turn angles are not included but a salt bridge is described. Using a PCA analysis, the energy wells can be observed and transitions are suggested for different conformations. The peptide is described as moving easily between basins with an energy barrier of $\approx k_B T$. There are multiple minima at 150K $<$ T $<$ 300K.

The Met-Enk paper by Garcia and Sanbonmatsu [53] shows the RA for five residues using an Assisted Model Building with Energy Refinement Parameter force field: 1994 version (AMBERP). AMBERP is explained in a paper by Cornell et. al. in [41]. To summarize, one solvated Met-Enk system at constant (N,V,T) with T=300K for 32 ns is simulated. Constant temperature MD with initial configuration equilibration temperatures of T = 275, 325, 375, and 450K was used with a replica exchange of temperatures. The Sanbonmatsu et. al. study [53] uses (N,V,T) in vacuum. Multiple minima are found from 150K to 300K indicating greater structural variablity at low temperatures.

The free energy landscape and dynamics of Met-Enk were also studied by Evans and Wales with parallel tempering MC and the discrete path sampling method. Using an implicit solvation potential, the Effective Energy Function: 1 (EEF1), and the force field CHARMM19 with an MC variant with minimization called basin-hopping was performed (as noted in Table 2.1). Steepest decent paths connect the minima and are stored in a database with a free energy disconnectivity graph. The relative free energies and occupation probabilities are compared, similar to previous methods discussed by Krivov and Karplus [91]. Specifically, the known rate constants between (groups of/a) minima are applied via a transition matrix representing the master equation to solve a vector of occupation probabilities. Perturbations of the closest paths to calculate and store new rate constants define an energy landscape. The Type II' β-turn is the lowest energy structure using the outlined technique and a rate of 3.1x10^9 s^{-1} at 298K was calculated. Thirty-eight groups of minima are found with in 90% of the population, the lowest at +2.5 kcal/mol.

130 ns Langevin dynamics trajectory of Met-Enk [92] samples basins with different dynamical descriptions depending on force field. CHARMM-27, AMBER-94, and CHARMM-19 trajectories show the peptide spending maximal time in basin 2. All force-fields differ in their charge distributions on the aliphetic groups and CHARMM-19. The AMBER-94 Met-Enk trajectory is strikingly different from that computed with AMBER-96. Additionally, these differences yield Gly-3 in different basins in the dihedral angle space. CHARMM-27 and AMBER-94 exclude Gly-3 from basins 1 and 6. OPLS-UA dynamics behaves very similar to that of CHARMM-27 with respect to Gly-3 basin populations. AMBER-96 shows Gly-3 in basins 1, 2, and 3. The OPLS-UA and OPLS-AA force fields exhibit similar behaviors, a Gly-3 preference is seen for basins 1, 5, and 7. It is interesting to note that PARM94 and PARM96 dihedral distributions from [53] are similar to AMBER 94 and AMBER 96 results.

2.5 Literature Consensus of Properties of Met-Enkephalin from Simulation

Conventional simulations show great flexibilty and lack of definite conformations in water [30, 32, 53, 56, 61, 84, 92–96]. The temperature of folding, T_f, is around 230K, the collapse transition is approximately 290K but ranges from 260K to 310K [58, 59] in various solvent models. Table 2.1 lists the details of the solvent models associated with T_f and T_θ.

The sampled backbone dihedral angle conformations show consistent β and α helices for all studies for Met-5 and Tyr-1. The Gly-2, Gly-3, Phe-4 (ϕ, ψ) configurations are varied and all show β-turns and alpha helices stablized by hydrogen bonds between different residues. The low potential energy state consistently found indicates a Type II' β-turn and the global minimum below +2.5 kcal/mol. Met-Enk is anticipated to be a weak folder and will have selectivities for the σ-type receptor [68].

Chapter 3

STMD Studies of Met-Enkephalin

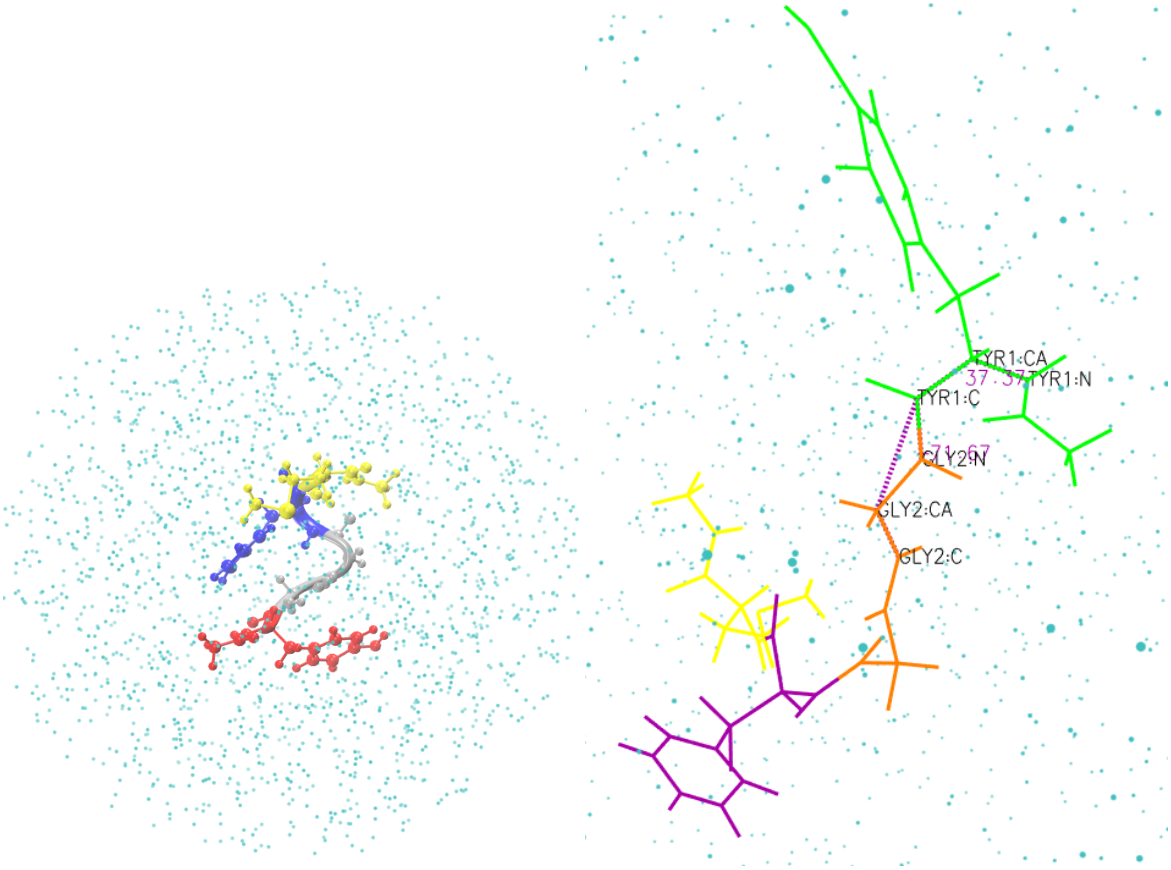
STMD studies were performed on the Met-Enk (YGGFM) peptide with 1917 water atoms. Calculations of average properties were compared to other simulations. Dihedral angle distributions reported by previous studies were compared and found to be consistent. Additional configurations were found that are not seen in previous studies. Properties of the system such as the $C(T)$ and T_θ are determined. The results presented here confirm that Met-Enk, a weak folding globular protein, exhibits characteristics of a disordered domain while retaining some features of a weak folder.

3.1 The Model of Met-Enkephalin in Water

The system consists of 2001 total atoms: 1917 CHARMM TIP3P water atoms make 639 waters. Eighty-four atoms compose a five amino-acid Met-Enk chain with methione terminal cap. All standard CHARMM input files are needed as well as extra STMD specific files for initialization at the beginning of an STMD run.

The CHARMM TIP3P water model differs from the original TIP3P model in that it has LJ parameters on the oxygen and hydrogen atoms. Due to these extra interaction sites, CHARMM TIP3P is suggested to be slower than TIP3P for some packages such as GROMACS [97]. Similarities between TIP3P and CHARMM TIP3P exist since both are three-site models and each interaction site holds the same point charge. The charges are taken from TIP3P for simplicity. The CHARMM TIP3P and original TIP3P models are noted to have the same freezing temperatures. For the purpose of this study, they are assumed identical in implementation and calculation of all properties, observables, and intrinsic/extrinsic values measured and reported in Chapters Three and Four. For the rest of the chapter, CHARMM TIP3P will be referred to as "the water model."

The force field used is the CHARMM22 version and the atom connections are determined by the corresponding CHARMM22 All-Hydrogen Topology File for Proteins [99].



A) Met-Enk with water droplet B) Met-Enk close-up

Figure 3·1: A water droplet with a restraining potential. 639 waters (1917 CHARMM TIP3P water atoms) and eighty-four atoms compose a five amino-acid Met-Enk chain with methione terminal cap using [98].

The energy terms include bonds, angles, Urey-Bradley, dihedrals, impropers, and residues. An energy correction map based on quantum mechanical calculations is applied (CMAP) [100, 101]. CMAP has improvements for conformational backbone properties which have the non-bonded LJ and non-bonded electric terms. The corresponding CMAP parameter file lists the relevant potential energy bonded distance changes with streching or contacting. The structural information of the system is stored in a Protein Structure File (PSF) for the Met-Enk with water system. A PSF was generated from NAMD [102] to create a 2001 atom water and protein system for CHARMM package with the STMD algorithm input.

For both Met-Enk and the water droplet, a boundary potential is employed and centered

at the origin. For CHARMM, the potentials are called with the keyword GEO from the charmm.inp file. The boundary potential prevents water from escaping into the vacuum, and keeps the center of mass of the peptide near the origin. The peptide is constrained with a harmonic potential, $V(r_{cm})_{harm} = 1/2K_b r_{cm}^2$, where K_b =10 kcal/mol/ Å^2 and r_{cm} is the Met-Enk center-of-mass.

Each water is constrained to the sphere by a potential depending on the distance, r_O, of the oxygen from the origin. The potential is zero for $r_O < r_{droff}$, where r_{droff} is a cutoff specifying the radius of the sphere, and rises as a quartic, $V(r_O)_{quartic} = a(r_O - r_{droff})^4$, for $r_O > droff$. In this study, a=10.2 kcal/mol/radian/radian and r_{droff}=15.5 Å.

The user ran a local supercomputing cluster version of modified CHARMM32b or CHARMM36b source code that remakes executables compiled with the required name "charmm.inp." This, along with other topology, parameter, and charmm setup files, contains information about the sequential exectution of all input parameters determined by the user.

The hydrogen-heavy atom bond lengths of Met-Enk and water are fixed during the dynamics calculations with the SHAKE [103] algorithm. Only a small deviation from the reference bond length parameter table values of 10^{-10} (no unit) is allowed. Since a high frequency of the hydrogen-heavy atom bond vibration would need a short time step, the bond lengths of all hydrogens in the total system of Met-Enk and water are fixed during the dynamics calculations. This speeds up the calculation.

STMD generates a generalized ensemble through canonical MD with forces scaled by the factor $\gamma(U)$. The kinetic temperature is set to T_0. Furthermore, ordinary canonical MD is used as part of the initial setup for STMD within the CHARMM package. In both cases, the velocity verlet (VV) integrator is employed. The frequency of the output of the gamma value and energy is set with the variable SPRNFRQ. In this work a timestep of 0.001 picoseconds was sufficient for VV integration. For these results, 1,250,000 steps/run is standard for a 24-hr wall-time job submit to the linux cluster batch system described in Section 1.7.

The most direct starting Met-Enk structure is a highly resolved crystal structure from the Protein Databank (PDB). Modifying and using an MD program to generate a well-minimized hybrid structure is the next step. There are PDB crystal structures in the PDB database that can be used as a starting structure – both were extensively modified to create the input files for this study. Since the PDB is unusable in its original form, a PSF file can be made from a PDB using various tools. For this process, first run all steps of regular MD equlibration without STMD. After standard tests for energy and temperature demonstrate appropriate heating, bonding pair matches should be checked for correct structural fluctuations. Other visual inspection methods and correlation measurements of average properties at each stage of regular MD equlibration should also be monitored. If possible to compute reliably, measurements of other intrinsic properties may give insight into the stability of the minimal structure found from quenching that may be used for stable equlibrium simulations.

When setting up the system using MD, the waters should be smoothy equilibrated to 300K. At long times in the traditional MD equlibration process, STMD may begin. Details of the procedure can be requested from the author of this manuscript.

The restart file is written at least once a run. Often the restart file is written at the end of the 24-hr run. The coordinates are written every 50 steps. A run is standard for a 24-hr wall-time job submit to the linux cluster batch system. The energy is written every 100 steps. The velocities can be written and the averages and root-mean-square fluctuations for the energy are calculated every 500 steps. The non-bonded list is regenerated every 10 steps to save on memory requirements. STMD's modification factor (f-value) for the DOS is 1.01 for this study.

The range of statistical temperatures is selected. An example of a setup for one set of results is as follows: the upper T_H is 700K and a lower T_L of 220K which correspond to the water model's freezing stability range. The offsets to the upper, SCTH, and lower, SCTH, ST are both set to 50. The histogram covers energy range (T_L + SCTL) to (T_H - SCTL). The energy ranges for STMD histogram statistics are set; SEMIN is -13000 kcal/mol and

SEMAX is -1000 kcal/mol. The binsize is 25. The kinetic temperature controlling the momenta, SETTEMP, T_0, is set to 300K. SHTol is the flatness tolerance set to 0.3. Note that the flatness tolerance can range from 0.2 to 0.3 which roughly correspond to 20% to 30% flatness. SPRNFRQ is set to 5000.

The constant temperature Nosè-Hoover thermostat [4] is utilized. The Nosè-Hoover thermostat as implemented in CHARMM is coupled to a one heat bath. The package exclusively uses the Nosè method for the temperature coupling in the canonical ensemble for greater accuracy. Other thermostats such as the Berendsen thermostat cannot correctly sample the canonical distribution and/or cannot be paired with the VV integrator. The original Hamiltonian for Nosè dynamics is

$$H = H_0 + HB = H_0(\frac{p}{s}, q) + \frac{P^2}{2Q} + (f+1)kTln(s) \tag{3.1}$$

where f is a degree of freedom of the physical system. The equations of motions defined by Eq. 3.1 are solved numerically to achieve a canonical ensemble MD simulation. A set of equations of motion which dispense with the time-scaling parameter s by transforming $p' = \frac{p}{s}$, $S = ln(s)$ and $dt' = \frac{dt}{s}$ where developed by Hoover and are accessible from the package. He defined the conserved quantity

$$H = H_0(p', q) + \frac{P^2}{2Q} + (f+1)kTS \tag{3.2}$$

The S in Eq. 3.2 is reset in the charmm.inp file under the STMD keyword with the acronym RSTN and a reference temperature of 300K. The magnitude of coupling, Q, from Eqs. 3.1 and 3.2 is set to a value of 1000. A convergence of the equations is achieved in ten recursive kinetic energy evaluations.

3.2 Simulation Results

A complete body of evidence for six temperature sets reveals consistent principles for meta-analysis that can be used as a tool for further data abstraction. These datasets sucessively drop T_L from 250K and keep T_H at or higher than 700K. These sets have T_L=250K,

220K, 190K, 150K, 110K, and 50K whilst 700K $\leq T_H \leq$ 1100K. Most of these datasets are not presented since the package becomes unstable or the results are incomplete for proper Stage 3 analysis. The collective figures in this chapter visually illustrate that the package attains ergodicity with sampling from the lowest to highest energies simulated. The primary difference in progressive lower boundary temperature figures for each chain-dynamics simulation set are seen in the $C(T)$ and T_θ as compared to previously reported values. The simulation results are organized as follows:

1. Basic checks for proper sampling using a histogram and energy vs time.

2. Verification of the collapse temperature, T_θ, using both the reweighted $C(T)$ and the Statistical Temperature, T_S.

3. Best results for the temperature-dependent, reweighted distributions of dihedral angles and the radius of gyration.

Validation of the theoretical principles of a generalized ensemble sampling of a small system are presented in Figure 3·2 to Figure 3·10. Since calculating some properties with more certainty than the current literature can require even a broader energy region exploration and longer simulation times (within the ms timescale) [7], the long-time stability and convergence of the STMD sampling is highlighted. The following figures visualize a collapse temperature range where the $C(T)$ has a peak and T_S has a minimum in the derivative.

A simulation with a T_L temperature at 190K to T_H at 700K can be run to tune the parameters and model setup as detailed in Figure 3·2. 4.9375 μs is the total simulation time for a converged simulation starting from Stage 1 until Stage 3. This does not include equlibration time for regular MD prior to turning on STMD within the CHARMM package. Figure 3·2 shows that the simulation manages to maintain relative flatness ($< 20\%$) in the histogram for high energy regions. Figure 3·2 shows that the individual run (in red upright hatches) is as flat as the total simulation (in green cross-hatches) and the production run histogram (in blue asterisks).

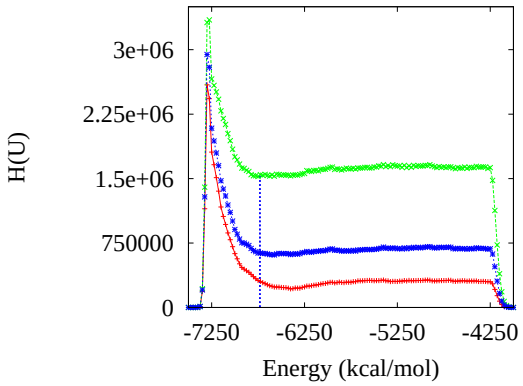


Figure 3·2: The Energy Histogram, $H(U)$, vs Energy Bin for T_L=190K. From bottom to top: the 172nd run in red upright hatches. The total histogram from Stage 1 is in green cross-hatches and the total histogram from the beginning of Stage 3 in blue asterisks. The blue dashed vertical line indicates the energy corresponding to T_θ=253K.

At the lowest energy regions, the Met-Enk simulation nearly breaks ergodicity. The sharp peak from -7400 kcal/mol to -6900 kcal/mol indicates sampling difficulty in a region below T_θ, see Figure 3·2. Thus the energy regions in the vicinity of T_θ are well sampled. The freezing temperature for the traditional TIP3P as well as this water model is 150K [104], corresponding to a ST of 0.5 and an energy of T_S=150K. The trapping was evident in the lowest energies indicating the start of a phase transition from liquid to an ice-like form. The freezing of the water determines the lowest T_L accessible with a flat histogram in these simulations.

Figure 3·3 shows Energy vs time for seventy-nine consecutive runs, which leads to histograms in Figure 3·2. The extensive Stage 3 energy exploration shows STMD to be an appropriate sampling algorithm for this temperature range.

The convergence of the simulation happens at the end of the simulation time of 4.9375 μs. The last energies in Figure 3·3 correspond to this timestep from the charmm output files. The low-energy trapping has been overcome when greater than 4.9375 μs of total simulation time has passed. A production run simulation time of 1 ns corresponds to 99999900 MD Steps as indicated along the x-axis in Figure 3·3. Slight dips from MD

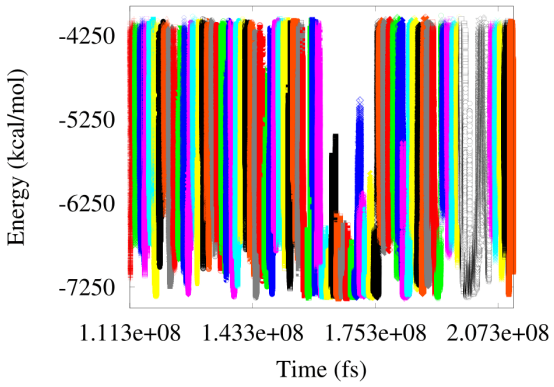


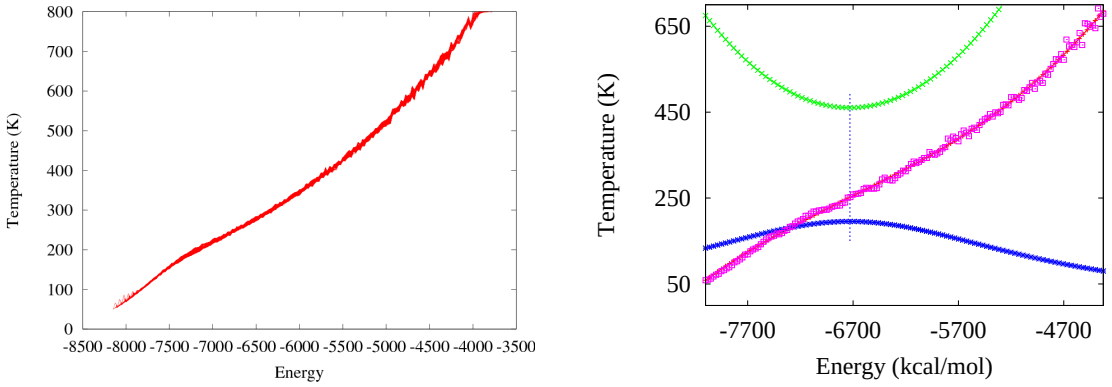
Figure 3·3: Energy vs time for 1.0 ns of the total production run for T_L=190K . 1 ns corresponds to 99999900 MD Steps. A point is printed every 1x10^{-15} seconds (fs).

Steps near $1.7 \times 10^{+12}$ steps correspond to the peaks near freezing outside the 20% flatness requirement maintained from the WL method in Figure 3·2a. The energy exploration over the total 4.94 μs trajectory shows sampling consistent with previous studies [36].

Previous runs for higher temperature ranges, where T_L=220K, 250K respectively, demonstrate flatter histogram sampling and more complete energy sampling for the whole MD trajectory. However, the T_L=190K to T_H=700K results reproduce $C(T)$, T_θ and other properties of the system more accurately within the limits of this implementation into CHARMM. Allowing the histogram to be less flat makes it possible to sample the energy region where the collapse transition takes place.

A collapse transition and temperature, T_θ, is usually identified by a peak in the canonical heat capacity, $C(T)$. The inverse of $\frac{dT_S(U)}{d(U)}$ is the microcanonical heat capacity, $C(U) = (\frac{dU}{dT(U)})$, which may be parametrically converted to a T-dependent approximation to $C(T)$ via $T_S(U)$. A peak in $C(T)$ is now seen to correspond to a minimum in the slope of $T_S(U)$. Thus collapse may be studied directly through the statistical temperature as well with $C(T)$.

When the energy range that includes the collapse region is sampled, the inverse of the derivative of the fit of T vs U does give a T_θ comparable to Evans and Wales (EW).



A) T vs U for T_L=50K, T_H=1100K B)T vs U with fits for T_L=50K, T_H=700K

Figure 3·4: A) Early Stage 1 STMD run after MD equilibration presenting the ST for T_L=50K to T_H=1100K. B) Temp. vs U. The ST curve in red is indistingushable from the cubic fit for the ST in magenta boxes. The green hatches are the derivative to the $T_S(U)$ fit represented as $\left[dT_S/\ dU\right]$. The blue cross hatches are $\left[dT_S/\ dU\right]^{-1}$. The peak of $\left[dT_S/\ dU\right]^{-1}$ is at -6730 kcal/mole and at 253K as indicated by the dotted vertical line.

Figure 3·4a displays the reproduceable T_θ from the early runs of a T_L=50K to T_H=1100K simulation. Figure 3·4a is the second twenty-four hour result beginning Stage 2 in the STMD algorithm within the CHARMM package. Stage 3 was not reached as the simulation became trapped at T_L.

Figure 3·4b indicates that going to T_L=50K reveals a region of minimal slope in $T_S(U)$, producing a peak in the inverse of the derivative at an energy corresponding to a temperature comparable to EW's T_θ value within 2 degrees K. The T_L=50K to T_H=700K simulations estimate a T_θ at 253K. The relation between the energy and temperature is determined with $T_S(U)$. The derivative of the magenta $T_S(U)$ curve is generated, which shows the minimum in the derivative of the ST.

To analyze T_S for the collapse transition and to determine T_θ, a cubic fit is constructed and represented as the magenta (central) curve that lies on top of the ST curve. The fit function is

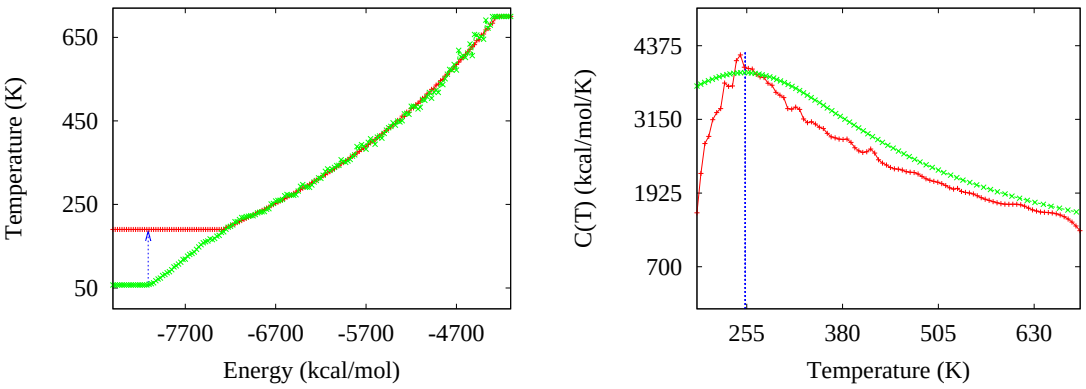
$$f(U) = T(U_1) + a\left[\frac{U+8100}{3775}\right] + b\left[\frac{U+8100}{3775}\right]^2 + c\left[\frac{U+8100}{3775}\right]^3 \tag{3.3}$$

where the ST at the lowest energy is set as $T(U_1)$=0.19574494. The coefficients to the cubic fit are a=2.24902, b=-1.97024, c=1.81002. These constants satisfy the requirement that the lower energy bound of the cubic fit, $f(U_1)$, is at T_L=50K. These equations are bounded by the lowest and highest energies simulated.

The derivative of the cubic fit is

$$\frac{df}{dU} = \frac{a}{3775} + \frac{2b}{3775}\left[\frac{U+8100}{3775}\right] + \frac{3c}{3775}\left[\frac{U+8100}{3775}\right]^2 \tag{3.4}$$

and is plotted in Figure 3·5 along with its inverse. The inverse expressed parametrically as a function of T using $T_S(U)$, an approximation to the canonical $C(T)$, is shown in Figure 3·6b.



A) ST difference for T_L=50K, 190K. B) C_v at T_L=190K. $C(T)$ fit at T_L=50K.

Figure 3·5: A) The $T_S(U)$ which exhibits a change in the ST from T_L=50K to 190K. The flat T_L=190K $T_S(U)$ is in red cross hatches. The curving $T_S(U)$ with T_L=50K is in green hatches. B) The average of the canonical $C(T)$ at T_L=190K using Eq. 3.5 in red crosses. The fit to the microcanonical $C(T)$ at T_L=50K in green hatches. The verticle blue dotted line indicates T_θ as the maximum for both peaks in the T_L=50K, 190K datasets. The colors of both lines in both sub-figures are the same for the same T_L.

Figure 3·5a and 3·5b show how using T_L=190K misses the collapse region in T_S. The canonical $C(T)$ in Figure 3·5b does reproduce the peak with T_L=190K, but the strong low temperature falloff is an artifact of the inadequate sampling range.

Figure 3·5 indicates there is no error in using the ST as a characteristic indicator of the canonical $C(T)$, but, choosing the correct temperature range is essential. The derivative of the ST for T_L=190K does not display a minimum corresponding to a peak in C(U) nor is it indicative of an S-loop in a stronger first order transition. For lower temperature ranges ($T_L \leq$ 190K), Stage 1 and 2 sampling reveal the characteristic signature in the ST collapse transition despite not achiving STMD equilibrium. Work by Bachmann [105] also gives a theory of phase transitions in terms of T_S.

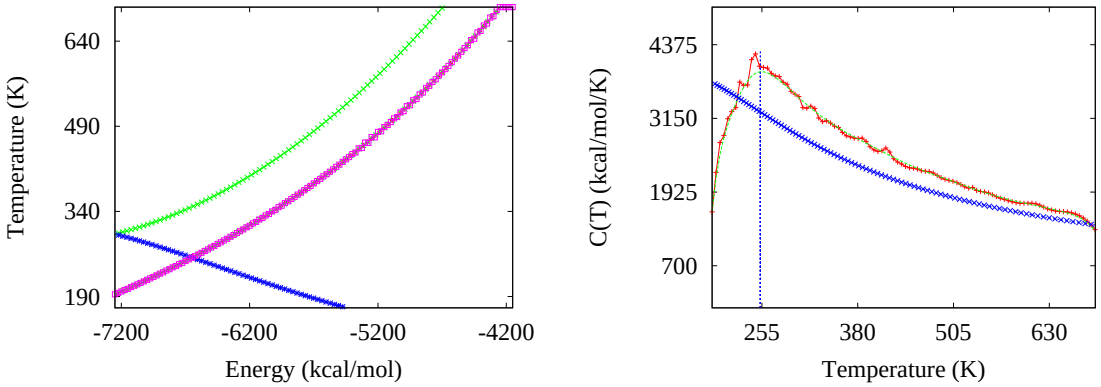
Choosing an appropriate T_L that samples well in any modified MD package may take some testing and tuning. Simulations with STMD within the CHARMM package at T_L=50K, 110K, and 150K became trapped in various Stages prior to STMD equlibration. For simulations with T_L above 190K, Stage 3 was reached and convergence of average properties and observables was found. The correct peak in the $C(T)$ was not verifiable for $T_L >$ 190K, but, more complete and flatter sampling results comparable to Figures 3·2 through 3·5 gave insight into choosing an intermediate T_L. A simulation with T_L=190K was able to complete to Stage 3.

The remainder of Chapter 3 is based upon runs with T_L=190K to a T_H=700K. Figure 3·5b shows the $C(T)$ comparison between the two most reliable datasets. The smooth curve representing the estimate of the microcanonical $C(T)$ from the T_L=50K dataset coincides with the rougher curve reflecting the canonical $C(T)$ with T_L=190K. The canonical $C(T)$ is calculated via a reweighting scheme using the weight discussed previously in Eq. 1.6. To get the canonical $C(T)$ in Figure 3·5b, average using Eq. 3.5 until convergence is verified.

$$C_v = \frac{<(\delta U)^2>}{k_B T^2} \tag{3.5}$$

The minor fluctuations after averaging 77 Stage 3 runs may indicate small sampling trappings. There is evidence from the T_L = 220K dataset which indicates the fluctuations will dissipate after longer simulations. A smoothing function can also achieve the same $C(T)$ curve if simulation time is limited, speeding up the process. The green line in Figure 3·5b has a bezièr fit to the rough peak. The peak in the canonical $C(T)$ is at 254K, the bezièr

fit peak is at 253K, and the peak of the estimate to the microcanonical $C(T)$ is at 253K. All curves can be interpreted to have excellent agreement within +/- 2K, which is a result unique to this study.



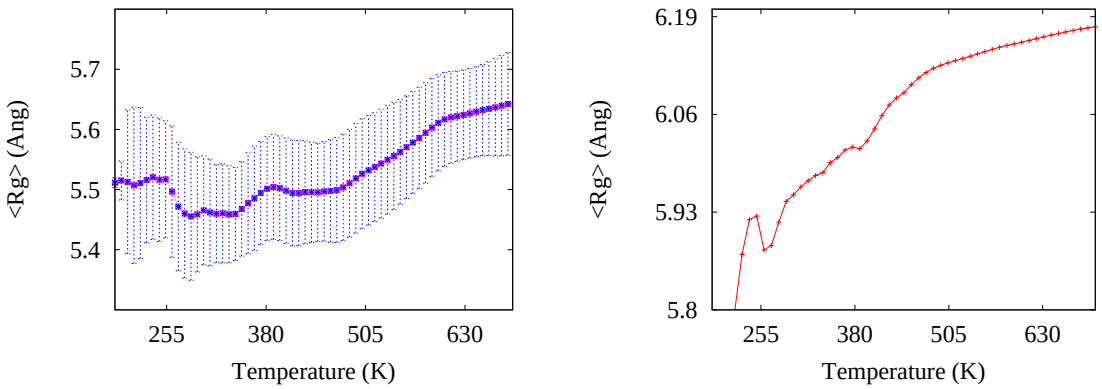
A) T vs U with fits for T_L=190K, T_H=700K B) $C(T)$ for T_L=190K, T_H=700K

Figure 3·6: A)Temp. vs U: $T_S(U)$ for T_L=190K in the green crosses. This ST line is covered by magenta diamonds representing the fit to $T_S(U)$. The derivative of the $T_S(U)$ fit is the upper curve in green hatches. $\left[dT_S/\ dU\right]^{-1}$ generates the line in blue asterisks. B) The red jagged line is the reweighted average $C(T)$ for T_L=190K, T_H=700K with a peak at 254K. The blue dashed line forms, from left to right, the micro-canonical $C(T)$ for the T_L=190K dataset.

Figure 3·6a and b confirm that T_S obtained with T_L=190K cannot directly give a peak in $C(U)$, while the reweighted $C(T)$ can. The inverse of the derivative of the fit to the ST curve has a peak at -7982 kcal/mol which is at 124K (not shown). A peak at a temperature below freezing for the water molecule could indicate the effect of the biomolecule on the freezing temperature. The protein is in a semi-frozen state and the conformations associated reflect this condition. These frozen-like states were explored in an effort to reproduce the lower temperature C(T). Other studies [59] report interesting features in the lower-temperature C(T) below 190K. Runs were done at 190K, 150K, 120K, 100K, 50K, and 0K in an attempt to push the CHARMM package with the STMD algorithm to the limits. As stated in the CHARMM documentation, the simulations below 150K became unstable [101] so it was anticipated that the STMD algorithm may stall. The C(T) features shown at different

lower-temperature simulations are beyond the scope of this dissertation.

When considering the total T_L=190K to T_H=700K simulation through all the Stages, the $C(T)$ peak fluctuated across the whole run from 230K to 260K and settled at around 255K. The shortest STMD Stage 3 time needed for consistent reproduction of the ST was two runs or approximately 0.9 μs. At infinite simulation time, the minor fluctuations in the $C(T)$ should become completely smooth using STMD. It is possible the minor variations are due to slight sampling trapping in difficult low-energy regions. Alternatively, the cause for trapping may be due to using the force-field outside of the temperature and energy range it was intended or pushing the sampling regions outside an accurate parameter set for the water model.



A) <Rg> vs. T for 53 runs for T_L = 190K. B) <Rg> vs. T for 32 runs for T_L = 220K.

Figure 3·7: A) The dark purple-pink solid line shapes the <Rg> vs. T for 53 runs with a T_L=190K found with CHARMM specific modules. The dotted purple error bars represent the standard deviation of the error at each temperature. B) <Rg> vs. T for 32 runs for T_L = 220K. Result found for MMTSB [106] with no error bars.

Figure 3·7a expresses the radius-of-gyration (R_g) using CHARMM at T_L=190K. Figure 3·7b is the MMTSB R_g result at T_L = 220K. The MMTSB module was used to calculate the R_g for T_L=220K.

Using two different calculation methods with different temperature regions, two different results were obtained. The MMTSB results are less accurate than the CHARMM

results due to low-level numerical simplifications and shortcuts used by MMTSB designers [106]. Also, the temperature range for Figure 3·7b is smaller and will show no indication of lower temperature trends $T_L < 220$K. The $< R_g >$ from both calculations and temperature ranges is consistent with Wodak's reported findings [52]. The fluctuations in the standard deviation of the errors are small but significant. The error bars show the increase in the R_g is slight indicating Met-Enk to be highly disordered at all temperatures sampled. For Figure 3·7b, the expected collapse signature is seen just above 255K in Figure 3·7. Work by Shen and Freed [56] and others [77, 96] support the weak foldability of Met-Enk.

3.3 Significant Configurations of Met-Enkephalin

The RA distributions for each amino acid in the neuropeptide Met-Enk (Tyr-1, Gly-2, Gly-3, Phe-4, Met-5) illustrate the ϕ and ψ angle control of the C_α-N and C_α-C distance. ϕ defines rotation about the C_α-N bond of the residue, and ψ defines rotation about the C_α-C bond of the same residue. This is calculated for all C_α residues along the backbone. The dihedral angles indicate the consecutive C_αs that pair into ϕ, ψ angles from a central rotation axes ranging from -180 to +180 degrees. The rotation about the central C_α-N and C_α-C bonding axes gives the different orientations. STMD within the CHARMM package samples all distributions of conformations found in the literature and includes more dense representations when compared to other results [53, 54, 59, 79, 107]. Configurations in Figures 3·8, 3·9 and 3·10 account for more densely sampled structures at all temperatures.

Main chain rotamers are angles measured as dihedral angles and may be indicated commonly with a ϕ, ψ, ω, or χ. Side-chain rotamers use the following sequence: α, β, γ, δ, ϵ, η symbols for successive rotamer angles moving along the side-chains moving away from the C_α atom of the backbone structure [108]. These bond descriptions and labeling systems are most relevant for main clusters of bend motifs [109]. The rotamer nomenclature used here is based on a trans conformation of Met-Enk peptide ω [79]. The ω angle is the angle of right-handed rotation about the C-N bond. If the C_α-C bond of the preceding residue is cis to the $N - C_\alpha$ bond, ω is zero. The planarity of the peptide bond usually restricts ω

to be 180° (the typical trans case) or 0 (the rare cis case) [110].

Bend motifs are relevant to physically observable structures seen by various experimental techniques. The most commonly observed are Type I and Type II′. Type II′ is described as an $\eta \rightarrow \alpha_R$ RA. In the rotamer nomenclature described using the main chain tagging system letters of "A," "B," etcetera, which are employed in these Figures, the Type II′ turn is found. The Type II′ turn will have the following main chain rotamer tags represented, which correspond to distributions in the RA Figures for Met-Enk: "B," "C," "A," and either "A" or "B." The other Type II′ rotamer indicator tags are: "B," "C," "B," and either "A" or "B." The superscripts to each rotamer region are meant to differentiate numeric tagging with the same main chain rotamer tags as indicated in the Figures for ease of use. The rotamer map presented here has the same definitions as Table 3 of the Carlacci study. Additionally, the same definitions of β turns are used in this work as for Carlacci's study [79]. Both these Type I and Type II turns are observed. The rotamer regions are shown in the RA figures for all temperatures simulated and have been cross-referenced with all stereochemical studies accessible and known to the author(s) of this dissertation.

Canonical averages, to find biologically relevant real temperatures, are obtained using reweighting techniques. In Figures 3·8, 3·9 and 3·10, the ln P(U) correction is applied to account for the possible 10% difference in the historgram flatness between the expected 20% flatness and any higher deviation from that percentage for T_L=190K. The ln P(U) correction is done to get the temperature dependence. Previous analysis of the T_L=220K to T_H=700K dataset as compared to the T_L=110K dataset, indicates that the differences between reweighted distributions are significant. There is little evidence of missing confirmations in the reweighted distributions for the T_L=110K dataset. The variations support the conclusion that the distributions of angles are consistently well sampled at 20% to 30% flatness for T_L=190K to T_H=700K for all temperatures.

Since Met-Enk is not a true helical protein there are conformations unique to a weak folder presented in these results as D, G, C, F, and H domains, illustrated in Figure 3·10. The domains, in combination, correspond to the following physical configurations. The

Residue	Group[a]	Angle (ϕ, ψ)
Tyr-1	D^a	(140 t 130, 180 t 153)
	$B0^b$	(175 t 170, 160 t 140)
	$B0^c$	(180 t 160, 153 t 120)
Gly-3	$B0^c$	(153 t 0, 180 t 160)
	AT	(0 t -50, -5 t -100)
	H	(180 t 160, 0 t -90)
	$B0^e$	(-120 t -180, -135 t -180)
Phe-4	$B0^c$	(180 t 160, 153 t 120)
	D^b	(75 t 40, -140 t -180)
	$B0^e$	(-120 t -180, -135 t -180)
Met-5	D^a	(80 t 40, 180 t 153)
	$B0^c$	(180 t 160, 153 t 100)
	AT	(-20 t -50, -5 t -85)
	G	(80 t -125, -100 t -155)
	D^b	(70 t 30, -120 t -180)
	$B0^e$	(-120 t -170,-155 t -180)
	$B1^e$	(-60 t -120,-160 t -180)

Table 3.1: Distribution details for the Group[a] and Angle (ϕ, ψ) are in Tables 4.1 and 3.2. Dihedral angles are listed which were not previously reported. These are unique to this study at T_θ=253K in comparison to MW's results. Note: "t" indicates the range of angles and representing the word "to". The superscript for Group[a] represents the rotamer region reading from left to right, top to bottom of the rotamer region. See Ch. 3.3 discussion for examples and work by Thornton, et. al. concerning stereochemical quality analysis [112].

"D" main chain rotamer tag is the domain that corresponds to a β-like turn indicated as a an ϵ RA [79, 109]. Thronton's work with Sibanda presents an image of the ϵ RA region [111]. Likewise, the "G" domain corresponds to a turn with character similar to an α or β turn that exhibits the angles necessary to fully characterize it as either motif. Figures of these regions are not shown in this work.

The populating of these rotamer regions indicates that many distributions are sampled in previously unsampled regions. Steric hinderance often causes forbidden regions. This study provides evidence that Met-Enk is not ever a well-folded protein in explict water. This chapter reports configurations that were difficult to sample in previous studies.

Joint Figures 3·8 and 3·9 arrange the reweighted RAs by increasing temperature. These are the results for a single run in Stage 3 for Gly-2. The distributions for Figure 3·8c to 3·8d at T_θ are shown so comparisons can be made to different literature values. Figure 3·9e distribution has a literature comparable temperature of 300K. After visual inspection of literature comparisions, a crude comparison shows STMD realized previously unknown conformations. Tables 2.1, 3.1, 3.2, and 4.1 reference these unique rotamer regions.

The relative amount of the distribution in the most probable bin is significant as it indicates a narrow distribution. The populations of distribution rotamer regions confirm this trend. The most probable group becomes less probable as the distribution broadens. A detailed study of the structural features expressed as x, y, z conformations was not done for regions not previously described in cited literature.

Specifically, each rotamer region population is divided from the most highly populated RAs at 0.1 to the lowest populated RAs of 0.1×10^{-9}. Each grouping is differentiated by 0.1×10^{-1}. Each 0.1×10^{-1} region is colored differently in the dihedral figures. Figures 3·8 and 3·9 color highlights Gly-2 populations for each 0.1×10^{-1} interval. When considering similar distributions for all five dihedral angles at all temperatures simulated, the 0.1×10^{-1} regions are most populated.

The visual effects of the combined shapes in the key show the most consistently sampled regions to be the blackish dots which indicate the most probable, or the 0.1 probabilities for (ϕ,ψ) angles. For the lowest 190K representation the 0.1×10^{-1} distribution has 25% of the total distribution concentrated around the darkest regions of Part A. The 216K RA's have 19% of conformations located in the 0.1×10^{-1} distribution. For T_θ at 252K, the orange filled upside down triangles represent 0.1×10^{-1} most probable distributions at 30% of the total density.

For each subfigure in Figures 3·8 and 3·9, the temperature dependent configurations at highest probability can be discussed. For the lowest 190K representation, the Gly-2 RA's at 0.1×10^{-1} within the probability distribution has 25% of the total distribution. The $B1^a$ and $B1^b$ rotamer regions of subfigure A have these distributions. The 216K

Gly-2 RA's occupy the $B1^a$ and $B1^b$ rotamer regions at 19% of conformations in the 0.1×10^{-1} distribution. The same β rotamer regions are represented for Gly-2 at 30% of the total density of the 0.1×10^{-1} most probable distributions at 257K. Above T_θ, the alpha lefthanded or "AL" or α_L turn is represented in the 298K distributions. RAs with 30% of all points at the 0.1×10^{-2} set of populations are in the "AL," $B1^a$ and $B1^b$ turns are populated for Gly-2.

This method predicts new 314K RAs for the highest probability rotamer regions include all previous regions and a $B0^d$ rotamer region. 40% of the distribution is in the 0.1×10^{-2} section. Angle distributions for 638K are in all the rotamer regions except the "AT" rotamer region. All of these configurations are highlighted in the 0.1×10^{-2} population set. This includes turns for the Type I, Type I′, Type II, and Type II′ turns. Type II′, which is expected to be the lowest energy configuration observed, is represented in this dataset.

Table 3.1 shows increased sampling using STMD. There is no direct and equivalent comparison to be made to Malevantes and Wodak's (MW's) study at 253K. However, MW's methodology and algorithmic techniques are most comparable when considering all studies which attempt to explore structural arguments.

In contrast, there is only one distribution at 253K that MW finds which this study does not, the D^a distribution. All other literature studies provided in the references are not directly comparable and were done with REMD, a modified generalized ensemble with different force scaling, or an MC variant. Furthermore, they were simulated at 300K so the distributions are not directly comparable to T_θ at 253K. The comparisons at 253K with MW's study are done with the knowledge that dihedral populations become more concentrated at lower energy minima as the temperature drops [113]. Thus, the 300K rotamer populations are assumed to be optimally populated outside low-energy regions. By process of elimination, comparisons to T_θ were deduced via observations of trends observed across temperatures in Figures 3·8 and 3·9.

All other studies report dihedral angles with about 30% less rotamer regions sampled. The Carlacci and Garcia studies are 35% less, MW reports 32% fewer regions, and Cukier

has no less than 49% loss in rotamer regions. These percentages are found using data reported in each study and extensive Tables (not shown here) mapping each rotamer region for each study in greater detail. The results of each study were taken from the most comparable datasets, many of which are not directly comparable but most closely approximate similar algorithms, simulation techniques, force-fields, system conditions, and methodology implementation for the entire literature on this system.

Traditional MD simulations consider a single temperature and report back dynamics as well as limited sampling of energies or other observables such as averaged groups of atoms with specific coordinates and bonding descriptions indicating structure. The structures reported are around 300K for biologically relevant and conformationally significant physical changes such as folding or collapse of the physical scaffold. An advantage to STMD is the ability to sample across a wide range of temperatures with one simulation trajectory. The limits for T_L and T_H are chosen specifically as relevant to the system of interest and other limitations on the measurements of the system. Early comparisons of structural indicators such as coordinated motions resulting in helicies, folds, turns, and coils near a point of greatest average heat change, or the peak of $C(T)$, shows STMD to have correctly estimated these structural transitions.

Figures 3·8 and 3·9 color codes the change in the probability in the dihedral angles for each 0.1 distribution grouping moving from lowest (top) to highest (bottom) temperatures. These 0.1 distribution groupings are divided into each rotamer region for Gly-2. The distributions which are most stable are Type'II and Type'I turns that are β [115]. These are low-energy structures that are well characterized in experimental studies. The next most stable regions are the α turns [111]. General helical regions are next most prominent at even higher energies. The highest energy regions are flexible and structurally disordered. As temperature increases past the collapse temperature, these disordered regions are well-sampled. The distributions of these regions are densely populated at 30% percent or higher (results not shown).

Gly-2 has been reported to be in a flexible region of the Met-Enk peptide. As the

temperature increases, more flexibility is found around these core structures (Figures 3·8 and 3·9A-H). The high temperature plots in Figure 3·9F shows a well-sampled total rotamer region. An in-depth analysis of these structures is not included here.

Figure 3·10 has the reweighted Ramachandran Angles (RA) for the 172nd run for the five amino acids at the T_θ for this study. The blackish dots indicate the most probable of sampled regions, or the 0.1 probablilty range for (ϕ,ψ) angles. Table 3.1 highlights the configurations using the same key as Figure 3·10 which are only found with this study.

To clarify Table 3.1, the heading Group^a with superscript a indicates rotamer subregions for one type of structural feature. Some turns occur multiple times, such as the β turns and disordered domains. When using similar labeling as Carlacci [79] it is noted that there are five subregions of β turns. If the rotamer key is read by "row" from top to bottom and left to right, then $B0^a$ to $B0^e$ are found successively from (-180, 180) to (180, -180). Figure 3·10 utilizes the same color key and rotamer region divisions differentiated by 0.1×10^{-1} as Figures 3·8 and 3·9. The populations of each distribution may be useful in elucidating β-factors for crystaline structures of Met-Enk [116].

Populations for configurations for Figure 3·10 at T_θ are as follows. Subfigure "A)" is Tyrosine-1 and has as the highest probability distribution the 0.1×10^{-2} group. The group has 33% of the total probability distribution. Subfigure "B)" is the Glycine-2 RAs. The 0.1×10^{-1} probability group has 30% of all conformations. Subfigure "C)" is of Glycine-3 and has the 0.1×10^{-1} probability set as the most densely populated with 30% of all RAs. Phenylalanine-4 shows 32% of all probability points at 0.1×10^{-2} in all relevant rotamer regions as subfigure "D)." The next subfigure "E)" has Methionine-5 with probable RAs at 32% of total conformations in the 0.1×10^{-2} set of populations. The final subfigure labeled "F)" is the rotamer map.

$\beta 0$ ranges in angle regions from $B0^a$ to $B0^d$. $\beta 1$ which consists of the $B1^a$ and $B1^d$ regions. The η RAs correspond to the domain "D" which has some disorder. The "D" domain also exhibits some β-like loose folding. A Folded domain is defined with the main chain rotamer region tag "F." The region exhibits δ-like and structures similar to right-

Residue	Group[a]	Angle (ϕ, ψ)
Tyr-1	D^b	(70 t 40, -140 t -180)
Phe-4	$B0^b$	(180 t 170, 180 t 153)
	$\beta 0^c$	(180 t 160, 153 t 120)
	D^b	(80 t 40, 180 t 153)
Met-5	D^a	(80 t 40, 180 t 153)
	$B0^b$	(180 t 160, 170 t 153)
	D^b	(70 t 30, -120 t -180)

Table 3.2: Dihedral angles not previously reported and unique to this study at T=300K. Other redundant distributions for the Group[a] and Angle (ϕ, ψ) are in the first column Table 4.1. These configurations can be viewed in Figures 4·2 and 4·3. Note: "t" indicates the range of angles and representing the word "to". The superscript for Group[a] represents the rotamer region reading from left to right, top to bottom of the rotamer region. The labeling corresponds to definitions by Thornton, et. al. concerning stereochemical quality [112].

handed γ secondary structure. The RA nomenclature for neighboring regions has an α Left Handed or "α LH" region, the α Right Handed or "RH" region. These regions are slightly differentiated from the α Turn or "α T" region, the Coil region as "C." This region has some η RA angle character but is verging on disordered. The η rotamer tag labeled "H" is also largely disordered. The "G" or γ region is structurally uncharacterized in the reference rotamer nomenclature. The labeling corresponds to IUPAC standards [108] and stereochemical determination analysis [111, 112, 114].

Biologically relevant temperatures of 300K for Met-Enk are represented in Figures 4·2 and 4·3 and the unique results for this study are in Table 3.2. When referencing the rotamer regions of the Carlacci study, all angles indicate previously unseen structures from experimental techniques.

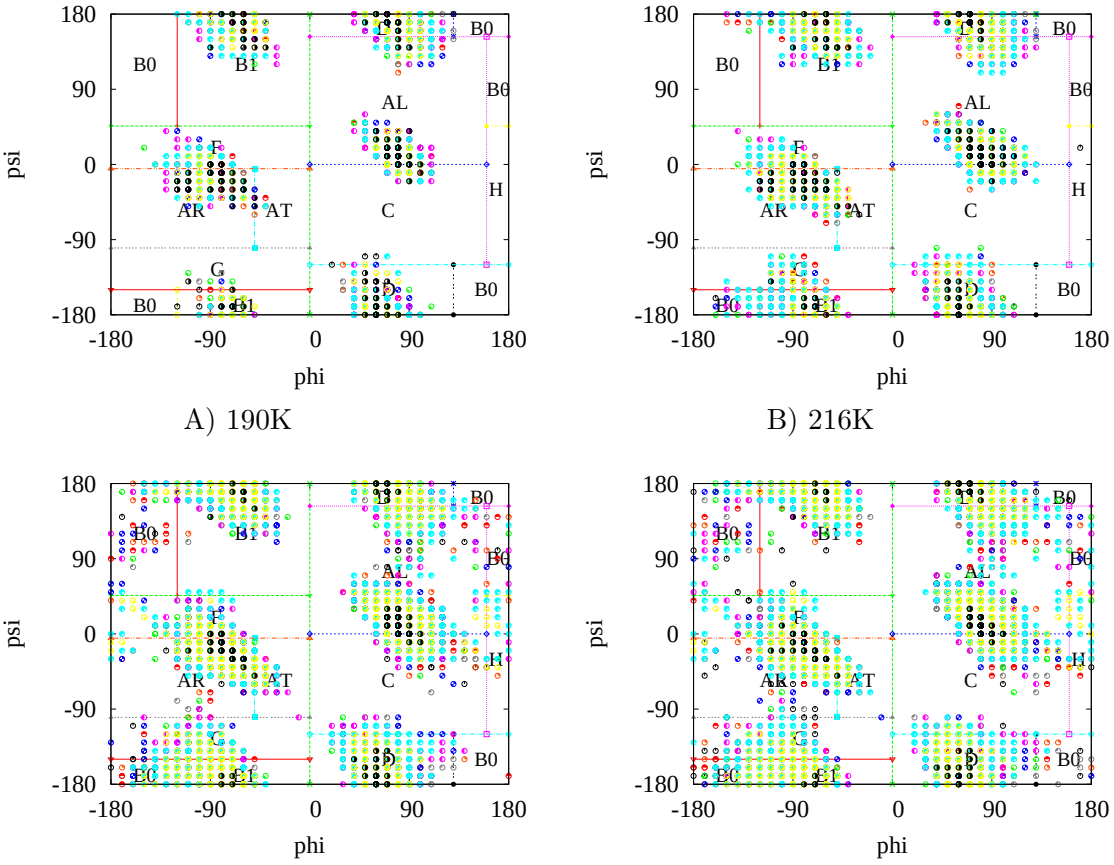
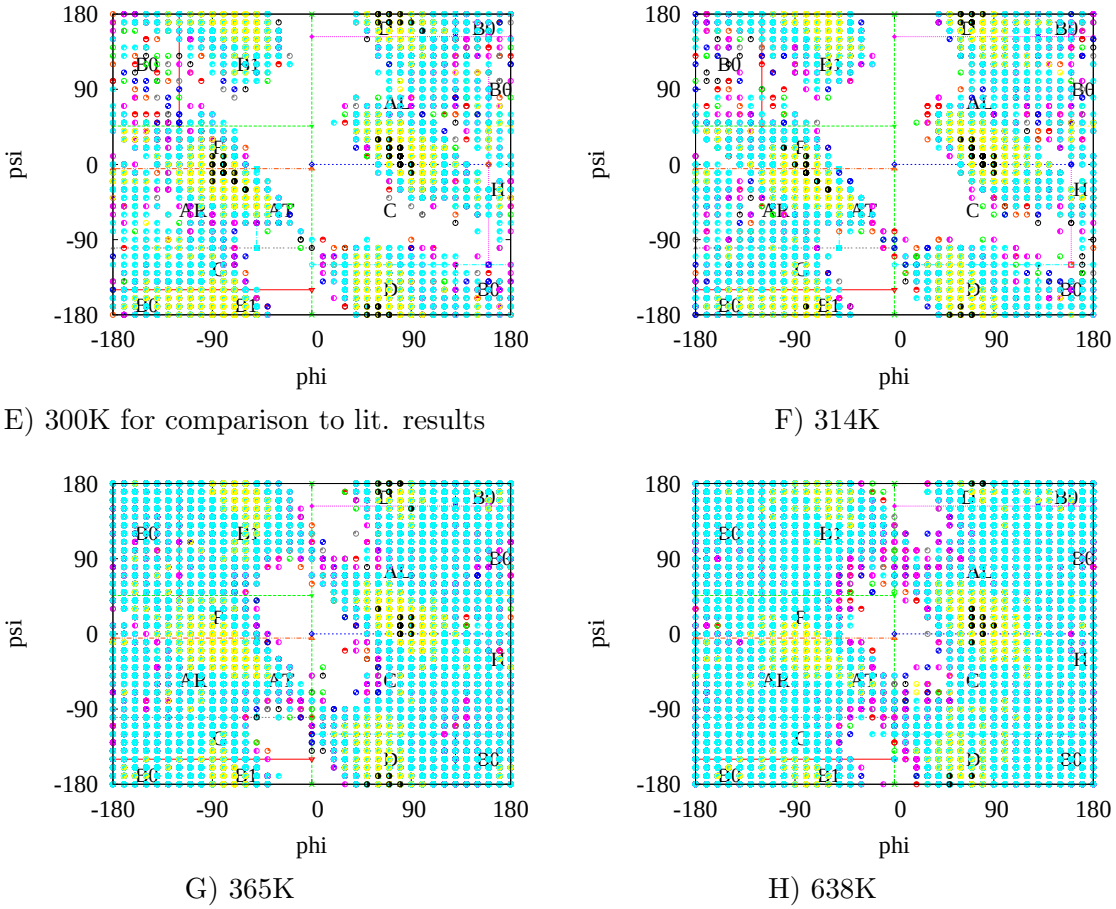


Figure 3·8: Gly-2 RAs from 190K to 257K. Each RA subfigure is spaced at an inverval across the lower half of the temperature range from T_L=190K to T_H=700K. Every distribution is separated into rotamer regions defined by Thronton and coworkers using stereochemical analysis [114]. The key is as follows: 0.1 is in red open diamonds, 0.1×10^1 is in orange filled upside down triangles, 0.1×10^2 is in gold brown upside down open triangles, 0.1×10^3 is in brown filled triangles, 0.1×10^4 is in light blue open triangles 0.1×10^5 is in magenta semi-diamonds, 0.1×10^6 is in blue open circles, 0.1×10^7 is in light green filled squares, 0.1×10^8 is in red open squares, 0.1×10^9 is in orange asteriks.



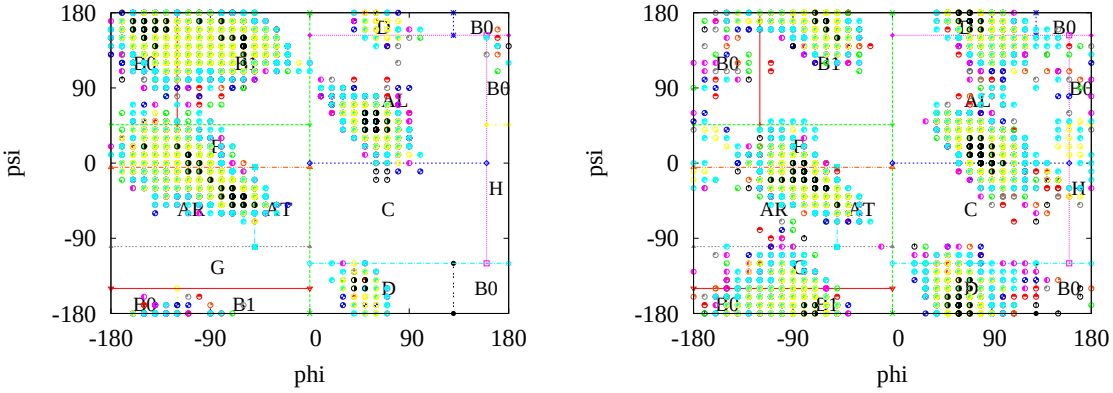
E) 300K for comparison to lit. results

F) 314K

G) 365K

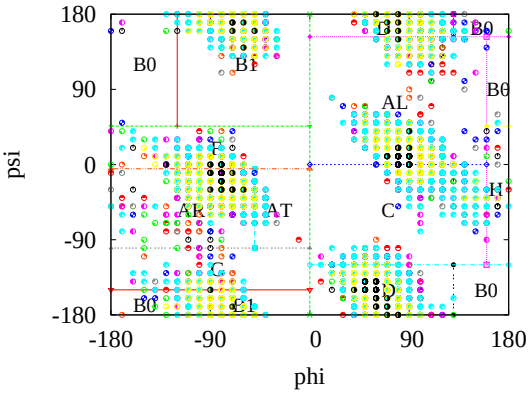
H) 638K

Figure 3·9: Gly-2 RAs from 300K to 638K. Each RA subfigure is spaced at an inverval across the upper half of the temperature range from T_L=190K to T_H=700K. Every distribution is separated into rotamer regions defined by Thronton and coworkers using stereochemical analysis [114]. The key is as follows: 0.1 is in red open diamonds, 0.1×10^1 is in orange filled upside down triangles, 0.1×10^2 is in gold brown upside down open triangles, 0.1×10^3 is in brown filled triangles, 0.1×10^4 is in light blue open triangles 0.1×10^5 is in magenta semi-diamonds, 0.1×10^6 is in blue open circles, 0.1×10^7 is in light green filled squares, 0.1×10^8 is in red open squares, 0.1×10^9 is in orange asteriks.

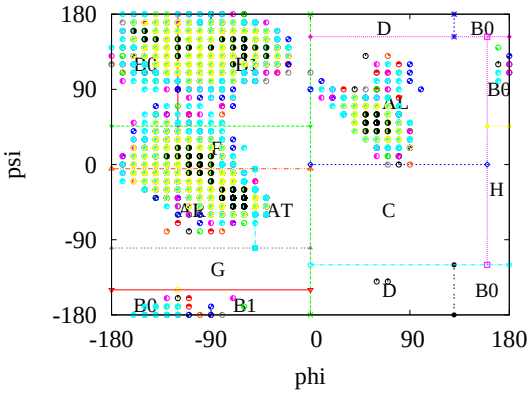


A) Tyr-1 253K

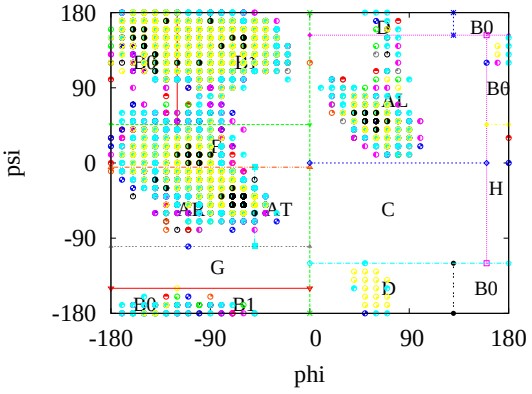
B) Gly-2 253K



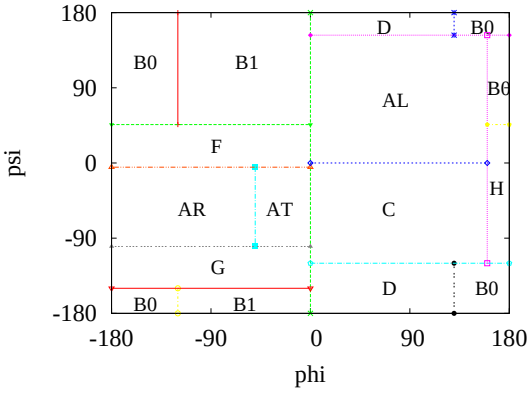
C) Gly-3 253K



D) Phe-4 253K



E) Met-5 253K



F) Rotamer regions

Figure 3·10: Met-Enk T_θ dihedral distributions. The key is as follows: 0.1 is in red open diamonds, 0.1×10^1 is in orange filled upside down triangles, 0.1×10^2 is in gold brown upside down open triangles, 0.1×10^3 is in brown filled triangles, 0.1×10^4 is in light blue open triangles 0.1×10^5 is in magenta semi-diamonds, 0.1×10^6 is in blue open circles, 0.1×10^7 is in light green filled squares, 0.1×10^8 is in red open squares, 0.1×10^9 is in orange asteriks.

Chapter 4

Comparison of Results to Literature

4.1 Discussion of Results

Findings in Chapter 3 will now be compared to the literature. The ramifications of other Met-Enk studies by many other scientists [53–55, 58, 59, 79, 115, 117–120] are also highlighted in this chapter.

4.2 Comparisons to Literature

STMD shows additional rotamer regions sampled. The "AL" or Alpha left-handed helix for Tyr-1 is a prominent addition; Figure 4·2a shows this left handed alpha helix. All amino acids show greater sampling of the beta-like disordered domains labled as "D" with STMD. One single-replica STMD simulation is seen to be competitive with 64-replica MRR, demonstrating the power of the algorithm. Results presented in Table 4.1 are consistent with CHARMM-19^{b2} results in Table 2.1. The Carlacci study [79] is relied upon for definitions of main chain rotamer regions that include definitions beyond the ϕ, ψ, angles to include the ω angles and other side-chain rotamers.

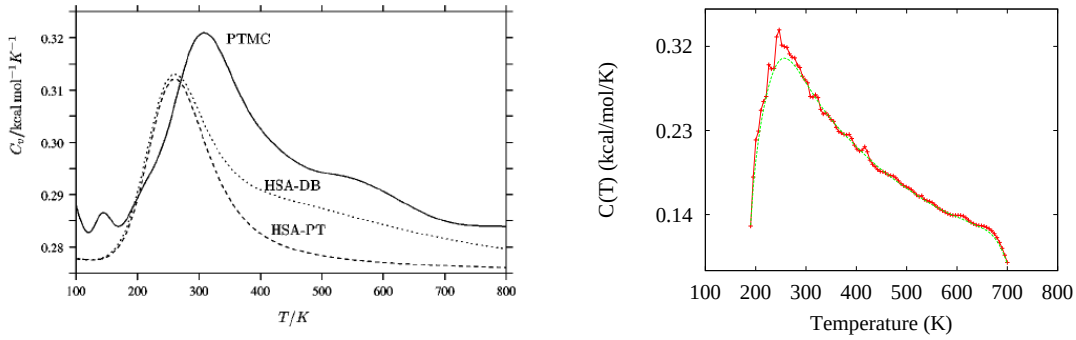
A REMD method is used to sample 16 temperature replicas from a system equlibrated at 300K in an NVT for 32 ns [53]. The approximately 600 explicit waters were equilbrated for 100 ps with periodic boundary conditions and a 9 Å cutoff in long-range force calculations. The sampled Tyr1 regions are the B0, B1, and AR, AT regions. The sampled Gly2 are B0, B1, F, AR, AT and AL, C, D, H regions. The Gly3 regions are AR, AT, F and B0, B1. The AL, C rotamer regions show interconverion similar to the Wodak paper distributions that can be found in [52]. The Phe4 results show B0, B1, and AR, AT detailed sampling; AL, C sampling is also shown. Met5 has B0, B1 sampling with AR, AT turns. Slight AL, C configurations are also found. These rotamer regions are found using the regions outlined by Carlacci [79] and based on work by previous researchers [108, 111, 112, 114].

4.2.1 Folding and Collapse Temperatures

The collapse temperature is found at 253K and is most consistent with the findings of Wales, et. al. using the CHARMM19^{b2} force field. Previously reported folding and collapse values are discussed in which ranged from 260K to 310K [58, 59]. Despite these studies using different force-fields and environments, the results presented in Chapter 3 fall in the range of these other studies. Table 2.1 reports collapse and folding temperatures of Met-Enk simulations [59].

The order parameter used to define the folding temperature is the fraction of native contacts. It is expected that the folding and collapse temperatures for small proteins that are good folders should be close [121, 122]. Met-Enk is not considered to be a good folder; it is relatively disordered and flexible. A high degree of frustration in the Met-Enk system indicates many mis-folding funnels in the energy landscape [123]. This indicates that the collapse and folding temperatures should occur at different temperatures, if they both occur at all. This work shows only a collapse temperature, suggesting that there is no folding temperature. Thus, the fraction of native contacts cannot be computed and described.

The Met-Enk vacuum simulation by Kim et. al. is in vacuum and quickly finds the native state and other highly flexible states close to the native state as compared to water or other solvent simulations [55]. Kim samples the complete conformational space very quickly. In test simulations, not reported here, a complete sampling of the vacuum Met-Enk system with one replica of STMD took 37,500 ps. The wall-time was less than 42 hours and only 13 hours dedicated CPU time which is exceedingly fast when compared to the same sampling for the simulations reported. The T_L=190K to T_H=700K simulations reported in Chapter 3.1 took approximately ten months of dedicated single CPU time per dataset using explicit water. These tests were made with the same hardware, software, implementation, and multiprocessing environment.



A) Wales' $C(T)$ for Met-Enk B) $C(T)$ for this study

Figure 4·1: A) The $C(T)$ for Evans and Wales study [59] with implicit solvent. B) The converged result using CHARMM with the STMD algorithm for Met-Enk using 639 explicit waters and T_L=190K.

4.2.2 Heat Capacity

The heat capacity is not directly comparable to any previous studies since the force-field, enhanced sampling algorithm, atom setup, simulation parameters, and water type are unique to this study. Crude comparisons can be made to another CHARMM simulation using the CHARMM19 force field. The Evans and Wales simulation uses a parallel-tempering (PTMC) and the harmonic superposition approximation (HSA). Figure 4·1 has good agreement, given that T_L limits STMD on the low temperature side, between the HSA-DB and STMD results. Evans and Wales show smoother average properties and compare to reported experimental peaks below 250K [59].

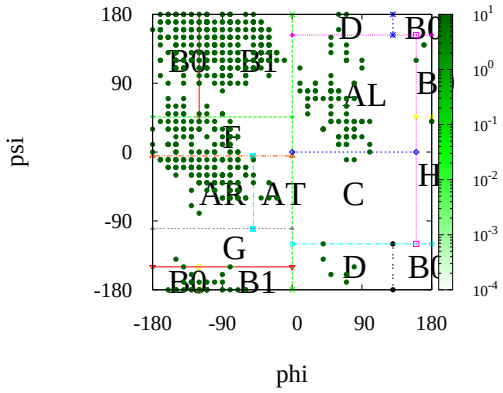
4.2.3 Dihedral Distributions

Chapter 2.4 paragraph two and paragraph ten in Section 2.4 introduce a study by Malevantes and Wodak (MW) [52]. These authors design a biased sampling algorithm with a potential designed to choose unpopulated or improbable conformations of Met-Enk. The CHARMM27 force field with 1000 explict TIP3P waters and truncated octahedron boundary conditions was simulated. It is unknown if CMAP corrections were considered [124]. After equlilbration, sixty-four replicas were equlibrated for 1 ns each. The last 0.5 ns

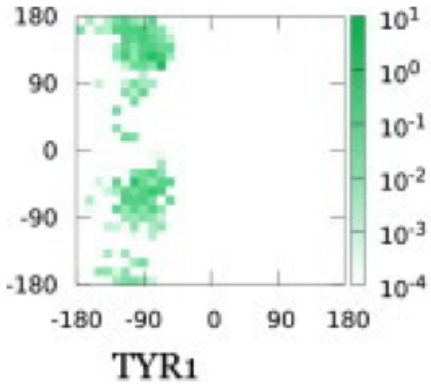
of each replica was used for analysis, yielding a total of 32 ns. Every 1 ps represents a snapshot for which the MRR weight was computed.

Table 4.1 compares MW's distributions and our results from Chapter 3. The dihedral structural features in Figures 3·8, 3·9, Figure 3·10, and Figures 4·2 and 4·3 show the range of temperatures that can be sampled from one simulation dataset where T_L=190K. MW used a constant temperature of 300K allowing a comparison to Chapter 3 data. Similarities between MW's simulations and these include explicit water with the TIP3P water model [99] modified for the CHARMM package.

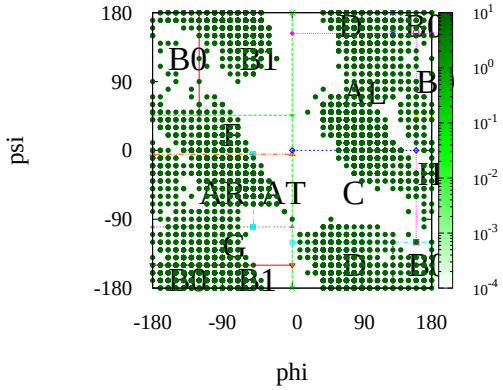
As previously stated, MW studied Met-Enk in explicit water at 300K with the Multiple Replica Repulsion (MRR) method, using 64 replicas. MW were able to sample configurations not found in a reference canonical MD run. The configurations they sampled can be found in Table 4.1.



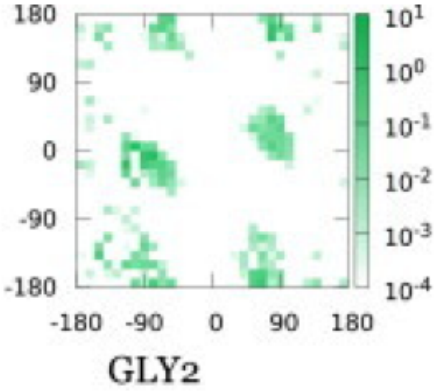
A) Tyr-1 at 301K



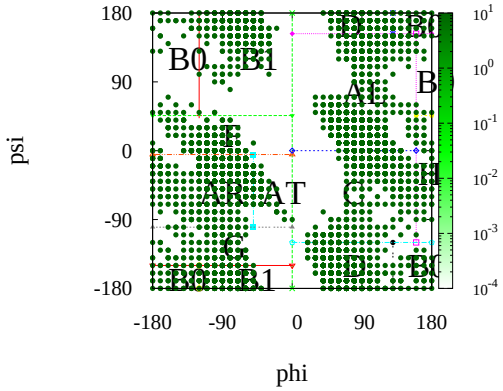
B) MW's Tyr-1 at 300K



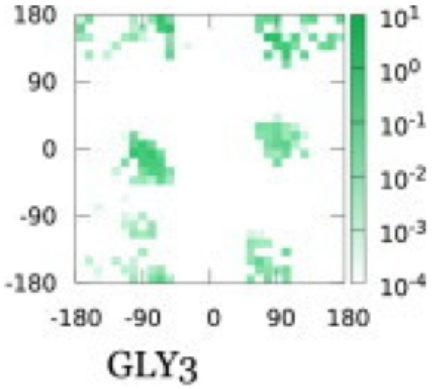
C) Gly-2 at 301K



D) MW's Gly-2 at 300K

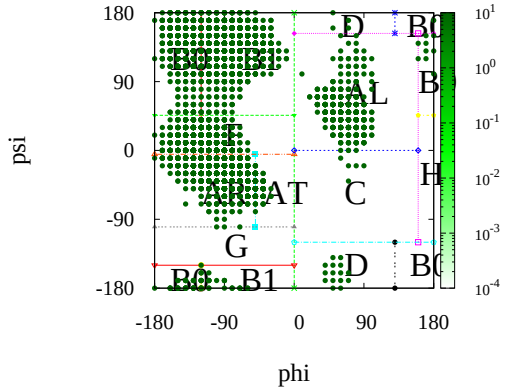


E) Gly-3 at 301K

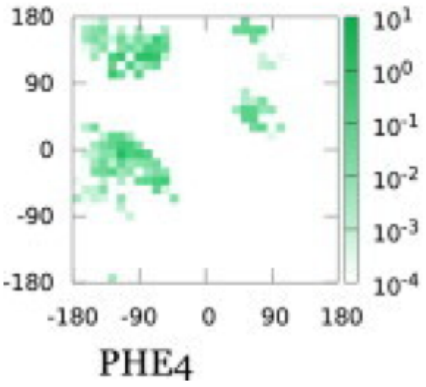


F) MW's Gly-3 at 300K

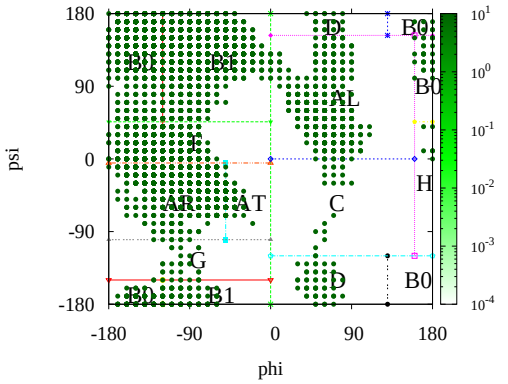
Figure 4·2: A) Wodak study comparison with first three amino acid dihedral angles.



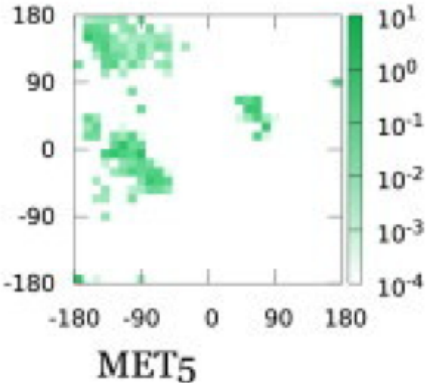
A) Phe-4 at 301K



B) MW's Phe-4 at 300K



C) Met-5 at 301K



D) MW's Met-5 at 300K

Figure 4·3: A) Wodak study comparison with last two amino acid dihedral angles.

	Current Met-Enk		MW	
	(ϕ,ψ)	$Turn$	(ϕ,ψ)	$Turn$
Tyr-1				
$B0^a$	(-120 t -180, 180 t 46)	β0	(-120 t -180, 180 t 46)	β0
$B1^a$	(-5 t -120, 180 t 46)	β1	(-50 t -120, 180 t 46)	β1
D^a	(140 t 130, 180 t 153)	Disord.		
F	(-70 t -180, 46 t -5)	Folded		
AL	(90 t 0, 153 t 0)	α LH		
$B0^c$	(170 t 170, 160 t 140)	β0		
AR	(-50 t -170, -5 t -70)	α RH	(-50 t -165, -10 t -100)	α RH
AT	(-25 t -50, -35 t -55)	α T		
C	(80 t 70, 0 t -10)	Coil		
G	(-145 t -155, -134 t -140)		(-90 t -155, -120 t -135)	
D^b	(70 t 40, -140 t -180)	Disord.		
$B0^e$	(-120 t -165, -135 t -180)	β0	(-120 t -165, -135 t -180)	β0
$B1^e$	(-60 t -120, -180 t -180)	β1	(-90 t -120, -150 t -180)	β1
Gly-2				
$B0^a$	(-120 t -180, 180 t 46)	β0	(-130 t -180, 180 t 46)	β0
$B1^a$	(-20 t -120, 180 t 46)	β1	(-50 t -110, 180 t 130)	β1
D^a	(130 t 40, 180 t 153)	Disord.	(90 t 115, 180 t 153)	Disord.
$B0^b$	(180 t 130, 180 t 153)	β0	(140 t 130, 180 t 153)	β0
F	(-60 t -180, 46 t -5)	Folded	(-45 t -120, 25 t -5)	Folded
AL	(160 t 15, 153 t 0)	α LH	(130 t 30, 55 t 0)	α LH
$B0^c$	(180 t 170, 153 t 46)	β0	(180 t 170, 153 t 130)	β0
AR	(-50 t -180, -5 t -100)	α RH	(-50 t -180, -5 t -60)	α RH
AT	(0 t -50, -5 t -90)	α T		
C	(160 t 10, 0 t -120)	Coil	(90 t 45, -90 t -120)	
H	(180 t 160, 46 t -120)	Helical	(180 t 175, -40 t -45)	Helical
G	(-55 t -180, -100 t -135)		(-85 t -155, -110 t -135)	
D^b	(160 t 10, -120 t -180)	Disord.	(90 t 30, -120 t -180)	
$B0^d$	(180 t 130, -120 t -180)	β0	(180 t 130, -130 t -180)	β0
$B0^e$	(-120 t -180, -135 t -180)	β0	(-120 t -165, -135 t -180)	β0
$B1^e$	(-45 t -120, -135 t -180)	β1	(-20 t -120, -135 t -180)	β1

	Current Met-Enk		MW	
	(ϕ,ψ)	$Turn$	(ϕ,ψ)	$Turn$
Gly-3				
$B0^a$	(-120 t -180, 180 t 46)	$\beta 0$	(-120 t -180, 180 t 120)	$\beta 0$
$B1^a$	(-20 t -120, 180 t 80)	$\beta 1$	(-40 t -120, 180 t 120)	$\beta 1$
D^a	(30 t 130, 180 t 153)	Disord.	(130 t 45, 180 t 153)	Disord.
F	(-50 t -180, 46 t -5)	Folded	(-45 t -130, 35 t -5)	Folded
AL	(160 t 30, 153 t 0)	α LH	(140 t 60, 45 t 0)	α LH
$B0^c$	(153 t 0, 180 t 160)	$\beta 0$	(153 t 0, 160 t 140)	$\beta 0$
AR	(-50 t -180, -5 t -100)	α RH	(-45 t -135, -5 t -100)	α RH
AT	(0 t -50, -5 t -100)	α T	(-45 t -50, -25 t -45)	α T
C	(160 t 20, 0 t -120)	Coil	(150 t 60, 0 t -25)	Coil
H	(180 t 160, 0 t -90)	Helical		
G	(-100 t -150, -60 t -180)		(-70 t -120, -100 t -130)	
D^b	(130 t 20, -120 t -180)	Disord.	(115 t 45, -120 t -180)	Disord.
$B0^d$	(180 t 130, -130 t -180)	$\beta 0$		
$B0^e$	(-120 t -180, -135 t -180)	$\beta 0$	(-50 t -170, -135 t -180)	$\beta 0$
$B1^e$	(-40 t -120, -135 t -180)	$\beta 1$	(-45 t -120, -155 t -180)	$\beta 1$
Phe-4				
$B0^a$	(-120 t -180, 180 t 46)	$\beta 0$	(-120 t -170, 180 t 120)	$\beta 0$
$B1^a$	(-5 t -120, 180 t 46)	$\beta 1$	(-40 t -120, 170 t 90)	$\beta 1$
D^a	(155 t 130, 180 t 153)	Disord.	(100 t 30, 180 t 153)	Disord.
$B0^b$	(180 t 170, 180 t 153)	$\beta 0$		
F	(-70 t -180, 46 t -5)	Folded	(-70 t -170, 45 t -5)	Folded
AL	(110 t 5, 153 t 0)	α LH	(110 t 40, 153 t 25)	α LH
$B0^c$	(180 t 160, 153 t 120)	$\beta 0$		
AR	(-50 t -180, -5 t -100)	α RH	(-40 t -180, -5 t -100)	α RH
AT	(-30 t -50, -25 t -80)	α T	(-30 t -50, -25 t -80)	α T
C	(90 t 60, 0 t -40)	Coil		
D^b	(75 t 40, -140 t -180)	Disord.		
$B0^e$	(-120 t -180, -135 t -180)	$\beta 0$	(-120 t -130, -170 t -180)	$\beta 0$

	Current Met-Enk		MW	
	(ϕ,ψ)	$Turn$	(ϕ,ψ)	$Turn$
Met-5				
$B0^a$	(-120 t -180, 180 t 46)	$\beta 0$	(-120 t -180, 180 t 95)	$\beta 0$
$B1^a$	(0 t -120, 180 t 46)	$\beta 1$	(-30 t -120, 180 t 46)	$\beta 1$
D^a	(80 t 40, 180 t 153)	Disord.		
$B0^b$	(180 t 160, 170 t 153)	$\beta 0$		
F	(-60 t -180, 46 t -5)	Folded	(-75 t -170, 46 t -5)	Folded
AL	(110 t 0, 153 t 0)	α LH	(100 t 50, 75 t 20)	α LH
$B0^c$	(180 t 160, 153 t 100)	$\beta 0$	(180 t 175, 90 t 80)	$\beta 0$
AR	(-50 t -180, -5 t -100)	α RH	(-160 t -50, -5 t -90)	α RH
AT	(-20 t -50, -5 t -85)	α T		
C	(90 t 50, 0 t -120)	Coil		
H	(180 t 170, 46 t 0)	Helical		

Table 4.1: Comparisons of literature values to values of the presented work to the MW's study [52]. The first column grouping shows STMD turns at T = 300K. The second column grouping is MW's findings. $B0^a$ to $B0^e$ are rotamer regions representing $\beta 0$ turns. D^a to D^b are disordered regions. All other regions are labled and can be cross-verified with Figures 4·2 and 4·3. Note: the superscript for Group^a represents the rotamer region reading from left to right, top to bottom of the rotamer region. The separator "t" indicates the range of angles and representing the word "to". Notation conforms with stereochemical quality analysis [112].

Table 4.1 shows that in general, the STMD study finds broader and more distributions than the MW study. The unique turns found with STMD are represented in all temperature regions. The turns found with STMD signify some low energy configurations that should have been found by other studies since some are stable β-hairpin structures and α turns, but they have only been seen with STMD, see Table 3.2 and Chapter 3 for detailed comparisons. The β turn is not confirmed for the (-180, -60) = (ϕ, ψ) region of Tyr-1 while the (-70,-40) configuration of Tyr-1 is not seen in these results as compared to an alternative analysis [53]. The α right handed turn is also extensively sampled with STMD for all residues, only Evans and Wales may sample more in this region than the STMD study.

The AL region is not sampled at all if STMD is not equlibrated or is in Stage 1 or 2. Once sampled in Stage 3, the "AL" region is seen over the whole energy range from T_L to T_H. Likewise, when selective short-time sampling in Stage 3 is analyzed, there is a low probability of seeing the "AL" regions within one run. The "AL" region is increasingly sampled as convergence for the average properties for different OP's is reached.

Other sterochemical regions only seen by STMD are disordered. This study finds many more disordered structures when compared to all other studies. There is a higher amount of disordered structures as temperature increases in STMD studies, with new regions showing in the slightly helical arrangements and disordered coils. The relevant disordered structures increase in distribution as temperature increases, leaving the overall distribution of all structures broadly distributed across all types of rotamer regions.

Residue	Group[a]	Angle (ϕ, ψ)
Tyr	D^b	(70 t 40, -140 t -180)
Phe-4	$B0^b$	(180 t 170, 180 t 153)
	$B0^c$	(180 t 160, 120 t 153)
	D^b	(75 t 40, -140 t -180)
Met-5	D^a	(80 t 40, 180 t 153)
	$B0^b$	(180 t 160, 170 t 153)
	D^b	(75 t 40, -140 t -180)

Table 4.2: Comparisons of literature values to values of the presented work to Garcia's paper [53], Culkier's study [54], Carlacci's results [79], and Wodak's results [52]. The grouping shows unique STMD turns at T = 300K. $B0^b$ and $B0^c$ are rotamer regions representing $\beta 0$ turns. D^a and D^b are disordered regions. Note: "t" indicates the range of angles and representing the word "to". The rotamer region key is in Table 3·10 part F and stereochemical quality analysis [112] The superscript for Group[a] represents the rotamer region reading from left to right, top to bottom of the rotamer region using IUPAC standards [108] and stereochemical quality [111, 112, 114].

Table 4.2 shows the rotamer regions unique to STMD results not seen in any comparison study besides MW's study. These studies attempt to present dihedral distributions as being completely sampled using different sampling techniques, force-fields, solvents, and methodologies. These comparisons are less-directly applicable, but, the ability of STMD to sample regions largely not seen in other techniques shows STMD to be sufficient to sample Met-Enk conformers well.

Su and Cukier studied turns for specific pairs of dihedral angles showing PCA first mode deep well configurations [54]. A comparison of Cukier's results may be possible by focusing on confined turns found in STMD at 300K. Figure 3·10 has the correct regions as cited in the paper by Sanbonmatsu and Garcia [53]. Figure 4 of the Garcia reference [53] shows in vacuum results, these results are in explicit solvent. The Carlacci study [79] shows some left handed alpha helix character, but the distribution is shifted and more disperse as compared to 4·2a. All other distributions are similar.

4.3 Computational Efficiency

Comparision benchmarking depends heavily on the averaging of several tests over exactly the same selections of a particular: programming language, compiler, compiler options, hardware configurations, multi-processing/programming environments. The algorithms being compared must all be implemented under the same conditions. Furthermore, parallelization of code complicates this analysis. Since STMD and the MMR methods were implemented under different conditions, precise comparisions cannot be made. Note that this is not a detailed comparison; rather it is an attempt to begin the precise efficency calculations. In the crude comparison attempted with this study, one single-replica STMD simulation may be competitive with 64-replica MRR as argued in the previous sections. Certainly, finding rotomers not found in previous studies is a plus for STMD.

When discussing regular MD, it is noted that these results do not sample the extensive structures that STMD can sample. The MD algorithm in a canonical ensemble [93] cannot report a representative dihedral distribution across a temperature range like STMD. MD at 300K only samples the narrow energy fluctuations corresponding to this temperature. A time step in femtoseconds is necessary in an MD simulation with an atomistic empirical force field. Bond stretching and bond bending are properly integrated by the equations of motions as the fastest motions of the system at a time-step of 10 to 15 seconds. More time would be needed for an ab initio calculation [125].

In general, RE combined with reweighting samples similar distributions to STMD, however, STMD samples many more rotamer regions. Parallelization may speed this up by a factor of 5 [55]. In other words, the primary advantage of RESTMD is it speeds up the determination of T_S. The large amount of time spent in Stages 1 and 2 will be shortened by RESTMD.

A rough comparision of the random walks in energy show less exploration than STMD [93] using results presented for a WL-type algorithm. In general, STMD should be more efficient than WL. This is done for model systems and should be comparable to metadynamics as

the STMD-metadynamics relationship was shown in the recent Vogel paper [126]. The computational efficiency of STMD may be comparable to accelerated dynamics routines [127].

A calculation of the Parallel Fraction of a Program using Amdahl's law [128] solved for the Fraction of the Program Code that can be made Parallel, P, is represented as

$$P = \frac{1}{2} \times \frac{t_2 - 1}{t_2} \tag{4.1}$$

where $P < 1$ and $t_2 = \frac{t_1}{t_n}$. t_1 is the time to execute the program with one-CPU and t_n is the time to execute with n CPUs. This second equation is a simple computational measure in terms of CPU time along with the parallelized hardware represented in Equation 4.2.

$$s(n) = \frac{t_1}{t_n} \tag{4.2}$$

where $s(n)$ is the speedup gained from applying n CPUs; again, t_1 is the time to execute the program with one-CPU and t_n is the time to execute with n CPUs.

The STMD simulation reported here is roughly equivalent to 1 canonical simulation covering the temperature range between T_L and T_H. In a compiler based and simple parallelization of 100 replicas, the speedup is estimated to be 5 to 20 times. Coding changes to the loop structures for STMD could increase this speedup by a small factor (2 to 10 times). This estimate comes from similar studies on the same system using different compilers, hardware, multi-processing environments, and programming environments. The t_1 estimate is 5 μs. The details of the speedup estimates for n CPU's are within the author's notes on test runs outside the scope of this study.

An unrefined method to begin standardization for STMD is to consider the time required to establish a flat energy histogram $P(U)$. While many collapse and expansion events have occured during this time; the ability to average a stable group of observables varies with the size of the system. The T_L=220K, and T_H=700K STMD simulations take 27,554 ps to reach production runs in Stage 3. There is 0.115 μs of simulation time needed for the 250K analysis to converge in average properties. During this simulation, numer-

ous semi-folding events occur over the entire energy range explored in a multi-canonical ensemble.

In another general comparison using similar starting conditions with the same hardware and similar setup system, test runs of a Met-Enk in vacuum take about eight runs (9,584 ps) to enter Stage 3; thirty five percent of the time needed to get the water all-atom STMD starting point. This is a rudimentary estimate of the time it takes for a simpler calculation without explicit water. The converged averages using STMD are useful for calculating the $C(T)$ and can require lengthy simulation times ($\geq$ 1 ms) [7] while the lowest-energy conformation and best comparision reweighted dihedral angles are found even after the first few runs ($\leq$3,750 ps). These time-scales show the incredible differences in simulation time needed and comparisions are difficult, even when changing the most simple of system details (the solvent).

Chapter 5

Summary and Future Directions

The first detailed application of STMD to an all atom representation of a biomolecule with explicit water is presented in this Dissertation. The estimate of the collapse temperature, $T_\theta = 253$ K, is in agreement with the paper of Evans and Wales. The same T_θ is obtained from the conventional method of identifying a peak in the heat capacity, C(T), and also from the minimum in the slope of the statistical temperature, $T_S(U)$.

With reweighing we have presented the first temperature dependent study of the RA distributions covering a broad temperature range. Rotamers found in the recent literature are reproduced but unique regions of RA were also found with STMD sampling. The new regions are mostly disordered domains and flexible regions which exhibit α and β turns at temperatures below T_θ. Bachmann et al have given an entropic theory of phase equilibrium in terms of $T_S(U)$.

Finding T_θ is just the beginning, and the complete theory may be applied in future work. It was found that freezing of the water limited the lower statistical temperature, T_L. Freezing in the presence of a biomolecule is a problem of considerable interest, and an appealing next topic. Stage 1 and 2, refining the statistical temperature estimate, took considerable CPU time. Replica exchange STMD can now be used and one of its advantages is greatly speeding up these steps. Applying it to Met-Enk is possible.

STMD obtains a flat distribution of the total energy. However, a flat distribution of a solute-related energy, ignoring the outer waters, may be more useful for bimolecular sampling. It is possible to modify STMD to get a flat distribution in any part of the energy, and Met-Enk would be a good target for the first application of "solute STMD."

Bibliography

[1] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller, "Equation of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, vol. 21, pp. 1087–1092, 1953.

[2] B. J. Alder and T. E. Wainwright, "Phase transition for a hard sphere system," *Journal of Chemical Physics*, vol. 27, p. 1208, 1957.

[3] M. P. Allen and Dominic J. Tildesley, "Unknown," in *Computer Simulation in Chemical Physics: Proceedings of the Nato Advanced Study Institute on New Perspectives in Computer Simulation in Chemical Physics*, Sept. 14-24, 1992.

[4] Shuichi Nose, "A unified formulation of the constant temperature molecular dynamics methods," *Journal of Chemical Physics*, vol. 81, no. 1, pp. 511–519, 1984.

[5] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Physical Review A (American Physical Society)*, vol. 31, no. 3, pp. 1695–1697, 1984.

[6] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[7] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, and Y. Shan, "Atomic-Level Characterization of the Structural Dynamics of Proteins," *Science*, vol. 330, pp. 341–346, 2010.

[8] Nobuyuki Nakajima, Haruki Nakamura, and Akinori Kidera, "Multicanonical Ensemble Generated by Molecular Dynamics Simulation for Enhanced Conformational Sampling of Peptides," *The Journal of Physical Chemistry B (ACS Publications)*, vol. 101, no. 5, pp. 817–824, 1997.

[9] D. J. Wales, "Energy landscapes: some new horizons," *Current Opinion in Structural Biology*, vol. 20, pp. 3–10, 2010.

[10] C. Levinthal, "Are there pathways for protein folding?," *Journal de Chimie Physique et de Physico-Chimie Biologique*, vol. 65, pp. 44–45, 1968.

[11] R. Zwanzig, A. Szabo, and B. Bagchi, "Levinthal's paradox," *Proceedings of the National Academy of Sciences, USA*, vol. 89, no. 1, pp. 20–22, 2010.

[12] J. W. Chen, P. Romero, V. N. Uversky, and A. K. Dunker, "Conservation of Intrinsic Disorder in Protein Domains and Families: I. A Database of Conserved Predicted Disordered Regions," *Journal of Proteome Research*, vol. 5, no. 4, pp. 879–887, 2006.

[13] M. Nakano, K. Kasai, K. Yoshida, T. Tanimoto, Y. Tamaki, and T. Tobita, "Conformation of the fowl protamine, galline, and its binding properties to dna," *Journal of Biochemistry*, vol. 105, pp. 133–137, 1989.

[14] V. N. Uversky, J. R. Gillespie, and A. L. Fink, "Why are natively unfolded proteins unstructured under physiologic conditions?," *Proteins: Structure, Function, and Bioinformatics*, vol. 41, no. 3, pp. 415–427, 2000.

[15] Jing Li, Hesam N. Motlagh, Carolyn Chakuroff, E. Brad Thompson, and Vincent J. Hilser, "Thermodynamic Dissection of the Intrinsically Disordered N-terminal Domain of Human Glucocorticoid Receptor," *The Journal of Biological Chemistry*, vol. 287, no. 32, pp. 26777–26787, 2012.

[16] Peter E. Wright and H. Jane Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm," *Journal de Chimie Physique et de Physico-Chimie Biologique*, vol. 293, no. 2, pp. 321–331, 1999.

[17] P. J. A. van Tilborg, F. A. A. Mulder, M. M. E. De Backer, M. Nair, E. C. Van Heerde, G. Folkers, P. T. van der Saag, Y. Karimi Nejad, R. Boelens, and R. Kaptein, "Millisecond to microsecond time scale dynamics of the retinoid X and retinoic acid receptor DNA-binding domains and dimeric complex formation," *Biochemistry*, vol. 38, pp. 1951–1956, 1999.

[18] Ken A. Dill and Justin L. MacCallum, "The Protein-Folding Problem, 50 Years On," *Science*, vol. 338, p. 1042, 2012.

[19] Agnes Tantos, Kyou Hoon Han, and Peter Tompa, "Intrinsic disorder in cell signal and gene transcription," *Molecular and Cellular Endocinology*, vol. 348, no. 2, pp. 457–465, 2012.

[20] Jessica H. Fong, Benjamin A. Shoemaker, Sergiy O. Barbuzynskiy, Michail Y. Lobanov, Oxana V. Galzitskaya, and Anna R. Panchenko, "Intrinsic disorder in Protein Interactions: Insights From a Comprehensive Structural Analysis," *PLoS Computational Biology*, vol. 5, no. 3, 2009.

[21] Tee Bordelon, Sarah K. Montegudo, Svetlana Pakhomova, Michael L. Oldham, and Marcia E. Newcomer, "A disorder to Order Transition Accompanies Catalysis in Retinaldehyde Dehydrogenase Type II," *Journal of Biological Chemistry*, vol. 279, no. 41, pp. 43085–43091, 2004.

[22] Bernd A. Berg and Tarik Celik, "New approach to spin-glass simulations," *Physical Review Letters*, vol. 69, pp. 2292–2295, 2012.

[23] Fugao Wang and D. P. Landau, "Efficient, Multiple-Range Random Walk Algorithm to Calculate the DOS," *Physical Review Letters*, vol. 86, no. 10, p. 2050, 2001.

[24] Koji Hukushima and Koji Nemoto, "Exchange Monte Carlo Method and Application to Spin Glass Simulations," *Journal of the Physical Society of Japan*, vol. 65, no. 6, pp. 1604–1608, 1996.

[25] Ayori Mitsutake, Yuji Sugita, and Yuko Okamoto, "Generalized-Ensemble Algorithms for Molecular Simulations of Biopolymers," *Biopolymers (Peptide Science)*, vol. 60, no. 2, pp. 96–123, 2001.

[26] Alan M. Ferrenberg and Robert H. Swendsen, "Optimized Monte Carlo data analysis," *Physical Review Letters*, vol. 63, no. 12, pp. 1195–1198, 1989.

[27] E. Marinari and G. Parisi, "Simulated tempering: A new monte carlo scheme," *Europhysical Letters*, vol. 19, no. 6, p. 451, 1992.

[28] Yuko Okamoto, "Protein folding simulations and structure predictions," *Computer Physics Communications*, vol. 142, pp. 55–63, 2001.

[29] Bernd A. Berg and Thomas Neuhaus, "Multicanonical algorithms for first order phase transitions," *Physical Letters B*, vol. 267, no. 2, pp. 249–253, 1991.

[30] Yuji Sugita and Yuko Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chemical Physical Letters*, vol. 314, no. 12, pp. 141–151, 1999.

[31] Jaegil Kim, Thomas Keyes, and John E. Straub, "Generalized Replica Exchange Method," *Journal of Chemical Physics*, vol. 132, p. 224107, 2010.

[32] Ulrich H. E. Hansmann, "Parallel tempering algorithm for conformational studies of biological molecules," *Chemical Physical Letters*, vol. 281, no. 13, pp. 140–150, 1997.

[33] David J. Earl and Michael W. Deem, "Parallel tempering: Theory, applications, and new perspectives," *Physical Chemistry Chemical Physics*, vol. 7, no. 23, pp. 3910–3916, 2005.

[34] H. Fukunishi, O. Watanabe, and S. Takada, "On the hamiltonian replica-exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction," *Journal of Chemical Physics*, vol. 116, no. 9058, 2002.

[35] P. Liu, B. Kim, R. A. Friesner, and B. J. Berne, "Replica-exchange with solute tempering: A method for sampling biological systems in explicit water," *Proceedings of the National Academy of Sciences, USA*, vol. 102, no. 39, p. 13749, 2005.

[36] Jaegil Kim, John E. Straub, and Thomas Keyes, "Statistical-Temperature Monte Carlo and Molecular Dynamics Algorithms," *Physical Review Letters*, vol. 97, no. 5, p. 050601, 2006.

[37] J. Kim, John E. Straub, and Thomas Keyes, "Statistical Temperature Molecular Dynamics: Application to course-grained β-Barrel-forming Protein Models," *Journal of Chemical Physics*, vol. 126, p. 135101, 2007.

[38] Jaegil Kim, John Straub, and Thomas Keyes, "Structure optimization and folding mechanisms of off-lattice protein models using statistical temperature molecular dynamics simulation: Statistical temperature annealing," *Physical Review E*, vol. 76, p. 011913, 2007.

[39] Juho S. Lintuvuori and Mark R. Wilson, "Statistical Temperature Molecular Dynamics Simulations applied to phase transitions in liquid transitions," *Journal of Chemical Physics*, vol. 132, p. 224902, 2010.

[40] Stefan A. Larrass, Laurel M. Pegram, Heather L. Gordon, and Stewart M. Rothstein, "Efficient generation of low-energy folded states of a model protein. II. Automated histogram filtering," *Journal of Chemical Physics*, vol. 119, p. 13149, 2003.

[41] W. D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *Journal of the American Chemical Society*, vol. 117, no. 19, pp. 5179–5197, 1995.

[42] Gregory H. Paine and Harold A. Scheraga, "Prediction of the native conformation of a polypeptide by a statistical-mechanical procedure. I. Backbone structure of enkephalin," *Biopolymers*, vol. 24, no. 8, pp. 1391–1436, 1985.

[43] Y. Isogai, G. Nemethy, and H. A. Scheraga, "Enkephalin: conformational analysis by means of empirical energy calculations," *Proceedings of the National Academy of Sciences, USA*, vol. 74, no. 2, pp. 414–418, 1977.

[44] F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, "Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids," *The Journal of Physical Chemistry*, vol. 79, no. 22, pp. 2361–2381, 1975.

[45] Z. Li and H. A. Scheraga, "Monte Carlo-minimization approach to the multiple-minima problem in protein folding," *Proceedings of the National Academy of Sciences, USA*, vol. 84, no. 19, pp. 6611–6615, 1987.

[46] M. Kinoshita, Y. Okamoto, and F. Hirata, "Solvent effects on conformational stability of peptides: RISM analyses," *Journal of Molecular Liqids*, vol. 90, pp. 195–204, 2001.

[47] M. Kinoshita, Y. Okamoto, and F. Hirata, "Solvation structure and stability of peptides in aqueous solutions analyzed by the reference interaction site model theory," *Journal of Chemical Physics*, vol. 107, pp. 1586–1599, 1997.

[48] A. Mitsutakea, M. Kinoshita, Y. Okamoto, and F. Hirata, "Multicanonical algorithm combined with the RISM theory for simulating peptides in aqueous solution," *Chemical Physical Letters*, vol. 339, no. 3-4, pp. 295–303, 2000.

[49] A. Mitsutakea, M. Kinoshita, Y. Okamoto, and F. Hirata, "Combination of the replica-exchange Monte Carlo method and the reference interaction site model theory for simulating a peptide molecule in aqueous solution," *Journal of Physical Chemistry B*, vol. 108, pp. 19002–19012, 2004.

[50] R. Radhakrishnan and B. L. Trout, *Order parameters in molecular simulations. Handbook of Materials Modeling.* Dordrecht: Kluwer Academic Publishers, 2005.

[51] Lewis C. Smeeton, Mark T. Oakley, and Roy L. Johnston, "Visualizing Energy Landscapes with Metric Disconnectivity Graphs," *Journal of Computational Chemistry*, 2014.

[52] A. Malevanets and S. J. Wodak, "Multiple replica repulsion technique for efficient conformational sampling of biological systems," *Biophysical Journal*, vol. 101, no. 4, pp. 951–960, 2011.

[53] K. Y. Sanbonmatsu and A. E. Garcia, "Structure of Met-Enkephalin in Explicit Aqueous Solution Using Replica Exchange Molecular Dynamics," *Proteins: Structure, Function, and Genetics*, vol. 46, no. 2, pp. 225–234, 2002.

[54] L. Su and Robert I. Cukier, "Hamiltonian and Distance Replica Exchange Method Studies of Met-Enkephalin," *Journal of Physical Chemistry B*, vol. 111, no. 42, pp. 12310–12321, 2007.

[55] Jae Gil Kim, Yoshifumi Fukunishi, and Haruki Nakamura, "Dynamical origin of enhanced conformational searches of Tsallis statistics sampling," *Journal of Chemical Physics*, vol. 121, no. 3, p. 1626, 2004.

[56] Min yi Shen and Karl F. Freed, "Long Time Dynamics of Met-Enkephalin: Comparison of Explicit and Implicit Solvent Models," *Biophysical Journal*, vol. 82, no. 4, pp. 1791–1808, 1993.

[57] U. H. E. Hansmann, Y. Okamoto, and J. N. Onuchic, "The folding funnel landscape for the peptide Met-enkephalin," *Proteins: Structure, Function, and Genetics*, vol. 34, pp. 472–483, 1999.

[58] U. H. E. Hansmann, M. Masuya, and Y. Okamoto, "Characteristic temperatures of folding of a small peptide," *Proceedings of the National Acadamy of Science USA*, vol. 94, no. 20, pp. 10652–10656, 1997.

[59] David A. Evans and David J. Wales, "The free energy landscape and dynamics of met-enkephalin," *Journal of Chemical Physics*, vol. 119, no. 18, p. 9947, 2003.

[60] D. J. Wales, "PATHSAMPLE: A Program for Refining and Analyzing Kinetic Transition Networks," July 29, 2011.

[61] U. H. E. Hansmann and Y. Okamoto, "Prediction of peptide conformation by multicanonical algorithm: New approach to the multiple-minima problem," *Journal of Computational Chemistry*, vol. 14, no. 11, pp. 1333–1338, 1993.

[62] J. G. Kim, Y. Fukunishi, A. Kidera, and H. Nakamura, "Determination of multicanonical weight based on a stochastic model of sampling dynamics," *Physical Review E*, vol. 68, no. 2, p. 021110, 2003.

[63] Jae-Suk Yang and Wooseop Kwak, "Application of Wang-Landau sampling to a protein model using SMMP," *Computer Physics Communications*, vol. 181, no. 1, pp. 99–104, 2010.

[64] Y. Okamoto, T. Kikuchi, and H. Kawai, "Prediction of Low-Energy Structures of Met-Enkephalin by Monte Carlo Simulated Annealing," *Chemistry Letters*, vol. 14, no. 3, pp. 1275–1278, 1992.

[65] Hagai Meirovitch, Eva Meirovitch, Andre G. Michel, and Maximiliano Vásquez, "A simple and effective procedure for conformational search of macromolecules: Application to Met- and Leu-enkephalin," *Journal of Physical Chemistry*, vol. 98, no. 25, pp. 6241–6243, 1994.

[66] U. H. E. Hansmann and J. Onuchic, "Thermodynamics and kinetics of folding of a small peptide," *Journal of Chemical Physics*, vol. 115, no. 3, pp. 1601–1606, 1981.

[67] J. E. Morley, N. E. Kay, G. F. Solomon, and N. P. Plotnikoff, "Neuropeptides: Conductors of the immune orchestra," *Life Sciences*, vol. 41, no. 5, pp. 527–544, 1987.

[68] J. Hughes, T. W. Smith, H. W. Kosterlitz, L. A. Fothergill, B. A. Morgan, and H. R. Morris, "Identification of two related pentapeptides from brain with potent opiate," *Nature*, vol. 258, pp. 577–579, 1975.

[69] Gavril W. Pasternak, "Multiple Morphine and Enkephalin Receptors and the Relief of Pain," *Journal of the American Medical Association*, vol. 259, no. 9, pp. 1362–1367, 1988.

[70] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry.* New York: W.H. Freeman, 5th ed., 2002.

[71] W. H. Graham, E. S. C. II, and R. P. Hicks, "Conformational analysis of Met-enkephalin in both aqueous solution and in the presence of sodium dodecyl sulfate micelles using multidimensional NMR and molecular modeling," *Biopolymers*, vol. 32, no. 12, pp. 1755–1764, 1992.

[72] D. Smith and J. F. Griffin, "Conformation of [Leu5]enkephalin from X-ray diffraction-features important for recognition at opiate receptor," *Science*, vol. 199, no. 4334, pp. 1214–1216, 1978.

[73] B. P. Roques, C. Garbay-Jaureguiberry, R. O. M. Anteunis, and A. K. Lala, "Conformation of Met-enkephalin determined by high field PMR spectroscopy," *Nature*, vol. 262, pp. 778–779, 1976.

[74] C. M. Deber and B. A. Behnam, "Role of membrane lipids in peptide hormone function: Binding of enkephalins to micelles," *Proceedings of the National Acadamy of Sciences USA*, vol. 81, no. 1, pp. 61–65, 1984.

[75] Victor J. Hruby, Lung Fa Kao, B. Montgomery Pettitt, and Martin Karplus, "The conformational properties of the delta-opioid peptide [D-PEN2, D-PEN5] enkephalin in aqueous-solution determined by NMR and energy minimization calculations," *Journal of the Americal Chemical Society*, vol. 110, no. 11, pp. 3351–3359, 1988.

[76] Y. Wang and K. Kuczera, "Molecular dynamics simulations of cyclic and linear DPDPE: influence of the disulfide bond on peptide flexibility," *Journal of Physical Chemistry*, vol. 100, no. 7, pp. 2555–2563, 1996.

[77] B. A. Behnam and C. M. Deber, "Evidence for a Folded Conformation of Methionine- and Leucine-Enkephalin in a Membrane Environment," *The Journal of Biological Chemistry*, vol. 259, no. 23, pp. 14935–14940, 1985.

[78] L. Zetta, A. D. Marco, G. Zannoni, and B. Cestaro, "Evidence for a folded structure of met-enkephalin in membrane mimetic systems: 1H-nmr studies in sodiumdodecylsulfate, lyso-phosphatidylcholine, and mixed lyso-phosphatidylcholine/sulfatide micelles," *Biopolymers*, vol. 25, no. 12, pp. 2315–2323, 1986.

[79] L. Carlacci, "Conformational analysis of [Met5]-enkephalin: solvation and ionization considerations," *Journal of Computational Aided Molecular Design*, vol. 12, no. 2, pp. 195–213, 1998.

[80] A. Aubry, N. Birlirakis, M. Sakarellos Daitsiotis, C. Sakarellos, and M. Marraud, "A crystal molecular conformation of leu-enkephalin related to the morphin molecule," *Biopolymers*, vol. 28, no. 1, pp. 27–40, 1989.

[81] Neeraj Misra, Onkar Prasad, and Leena Sinha, "Vibrational dynamics of morphine in relation to Leu5-and Met5-enkephalins," *Indian Journal of Biochemistry and Biophysics*, vol. 43, pp. 173–181, 2006.

[82] F. Shan; Y. Xia, N. Wang, J. Meng, C. Lu, Y. Meng, and N. P. Plotnikoff, "Functional modulation of the pathway between dendritic cells (DCs) and CD4+T cells by the neuropeptide: Methionine enkephalin (MENK)," *Peptides*, vol. 32, no. 5, pp. 929–937, 2011.

[83] U. Hansmann and Y. Okamoto, "Generalized-ensemble Monte Carlo method for systems with rough energy landscape," *Physical Review E*, vol. 56, no. 2, pp. 2228–2234, 1997.

[84] B. G. Nielsen, M. Ø. Jensen, and H. G. Bohr, "The probability distribution of sidechain conformations in [Leu] and [Met]enkephalin determines the potency and selectivity to mu and delta opiate receptors," *Biopolymers*, vol. 71, no. 5, pp. 577–592, 2003.

[85] Jacek Dudowicz, Karl F. Freed, and M.-y. Shen, "Hydration structure of metenkephalin: A molecular dynamics study," *Journal of Chemical Physics*, vol. 118, no. 4, p. 1989, 2003.

[86] Toshimasa Ishida, Shigetaka Yoneda, Mitsunobu Doi, Masatoshi Inoue, and Kunihiro Kitamura, "Molecular-dynamics simulations of [Met5]- and [d-ala2,met5]-enkephalins," *Biochemical Journal*, vol. 255, pp. 621–628, 1988.

[87] S. Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth, "Hybrid monte carlo," *Physical Letters B*, vol. 195, no. 2, p. 216, 1987.

[88] G. Parisi and Y. S. Wu *Scientia Sinica*, vol. 24, p. 483, 1981.

[89] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, "A Second Generation Force Feld for the Simulation of Proteins, Nucleic Acids, and Organic Molecules," *Journal of the Americal Chemical Society*, vol. 117, no. 19, pp. 5179–5197, 1995.

[90] J. W. Ponder, S. Rubenstein, C. Kudrot, S. Huston, M. Dudek, Y. Kong, R. Hart, M. Hodsdon, R. Pappu, W. Mooij, and G. Loeffler, "Tinker: Software Tools for Molecular Design, Version 3.7, Washington University," 1999.

[91] Sergei V. Krivov and Martin Karplus, "Free energy disconnectivity graphs: Application to peptide models," *Journal of Chemical Physics*, vol. 117, no. 23, p. 10894, 2002.

[92] Muhammad H. Zaman, Min Yi Shen, R. Stephen Berry, and Karl F. Freed, "Computer simulation of Met-Enkephalin Using Explicit Atom and United Atom Potentials: Similarities, Differences, and Suggestions for Improvement," *Journal of Physical Chemistry B*, vol. 107, no. 7, pp. 1685–1691, 2003.

[93] U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, "Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble," *Chemical Physics Letters*, vol. 259, no. 3, pp. 321–330, 1996.

[94] David van der Spoel and Herman J. C. Berendsen, "Molecular Dynamics Simulations of Leu-Enkephalin in water and DMSO," *Biophysical Journal*, vol. 72, no. 5, pp. 2032–2041, 1997.

[95] Mahalaxmi Aburi and Paul E. Smith, "A conformational analysis of leucine enkephalin as a function of pH," *Biopolymers*, vol. 64, no. 4, pp. 177–188, 2002.

[96] Dominic J. Tildesley, "Dmitry Nerukh; George Karvounis; and Robert C. Glen," in *Dynamic Complexity of Chaotic Transitions in High-Dimensional Classical Dynamics: Leu-Enkephalin Folding*, CompLife 2006, September, 2006.

[97] Pär Bjelkmar, B. H. Per Larsson, Michel A. Cuendet, and Erik Lindahl, "Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models," *J. Chem. Theory Comput.*, vol. 6, no. 2, pp. 459–466, 2010.

[98] W. Humphrey, A. Dalke, and K. Schulten, "VMD - Visual Molecular Dynamics," *Journal of Molecular Graphics*, vol. 14, pp. 33–38, 1996.

[99] Alexander D. Mackerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote,

J. Straub, M. Watanabe, J. Wiórkiewicz Kuczera, D. Yin, and M. Karplus, "All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins," *Journal of Physical Chemistry B*, vol. 102, pp. 3586–3616, 1998.

[100] Alexander D. Mackerell Jr., Michael Feig, and Charles L. Brooks III, "Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations," *Journal of Computational Chemistry*, vol. 25, no. 11, pp. 1400–1415, 2004.

[101] B. R. Brooks, C. L. Brooks III, A. D. Mac Kerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "CHARMM: The Biomolecular Simulation Program," *Journal of Computational Chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.

[102] James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Kaus Schulten, "Scalable molecular dynamics with NAMD," *Journal of Computational Chemistry*, vol. 26, pp. 1781–1802, 2005.

[103] J.-P. Ryckaert, Giovanni Ciccotti, and Herman J. Berendsen, "Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes," *Journal Of Computational Physics*, vol. 23, pp. 321–341, 1977.

[104] Carlos Vega and Jose L. F. Abascal, "Simulating water with rigid non-polarizable models: a general perspective," *Physical Chemistry Chemical Physics*, vol. 13, no. 44, pp. 19643–20022, 2011.

[105] Michael Bachmann, *Thermodynamics and Statistical Mechanics of Macromolecular Systems*. Cambridge CB2 8BS, United Kingdom: Cambridge University Press, 2014.

[106] Michael Feig, John Karanicolas, and Charles L. Brooks III, "MMTSB Tool Set: Enhanced Sampling and Multiscale Modeling Methods for Applications in Structural Biology," *Journal of Molecular Graphics and Modeling*, vol. 22, no. none, pp. 377–395, 2004.

[107] Xianmei Cai and Chhabil Dass, "Structural characterization of methionine and leucine enkephalins by hydrogen/deuterium exchange and electrospray ionization tandem mass spectrometry," *Rapid Communications In Mass Spectrometry*, vol. 19, no. 1, pp. 1–8, 2005.

[108] American Chemical Society, "IUPAC-IUB, Commission on Biochemical Nomenclature, Abbreviations, and Symbols for the Description of the Conformation of Polypeptide Chains," *Biochemical Journal*, vol. 121, no. 4, pp. 577–585, 1971.

[109] V. A. Efimov, *IAzyk ormuri v sinkhronnom i istoricheskom osveshchenii.* Moskva: Izd-vo "Nauka," Glav. red. vostochnoi lit-ry, 1986.

[110] Pierrick Craveur, Agnel Praveen Joseph, Pierre Poulain, Alexandre G. de Brevern, and Joseph Rebehmed, "Cis-trans isomerization of omega dihedrals in proteins," *Amino Acids*, vol. 45, no. 279-289, 2013.

[111] Bancinyane Lynn Sibanda, Tom L. Blundell, and Janet M. Thornton, "Classification of β-Hairpins in Protein Structures: A Systematic Classification with Applications to Modelling by Homology, Electron Density Fitting and Protein Engineering," *Journal of Molecular Biology*, vol. 206, no. 759–777, 1989.

[112] Anne Louise Morris, Malcolm W. MacArthur, E. Gail Hutchinson, and Janet M. Thornton, "Stereochemical Quality of Protein Structure Coordinates," *PROTEINS: Structure, Function, and Genetics*, vol. 12, pp. 345–364, 1992.

[113] John Kuriyan, Boyana Konforti, and David Wemmer, *The Molecules of Life: Physical and Chemical Principles.* Secaucus, NJ, U.S.A.: Garland Science, 2012.

[114] R. A. Laskowski, "PROCHECK: a program to check the stereochemical quality of protein structures," *Journal of Applied Crystallography*, vol. 26, pp. 283–291, 1993.

[115] K. Vengadesan and N. Gautham, "Conformational studies on enkephalins using the MOLS technique," *Biopolymers*, vol. 74, no. 6, pp. 476–494, 2004.

[116] L. Zhan, J. Z. Y. Chen, and W.-K. Liu, "Conformational study of met-enkephalin based on the ECEPP force fields," *Biophysical Journal*, vol. 91, pp. 2399–2404, 2006.

[117] C. Junghans and U. H. E. Hansmann, "Numerical Comparison of Wang-Landau Sampling and Parallel Tempering for Met-enkphalin," *International Journal of Modern Physics C*, vol. 17, no. 6, pp. 817–824, 2006.

[118] Christopher D. Snow, Eric J. Sorin, Young Min Rhee, and Vijay S. Pande, "How Well Can Simulation Predict Protein Folding Kinetics and Thermodynamics?," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 34, pp. 43–69, 2005.

[119] Elmar Bittner, Andreas Nubaumer, and Wolfhard Janke, "Make Life Simple: Unleash the Full Power of the Parallel Tempering Algorithm," *Physical Review Letters*, vol. 101, pp. 130603–1, 2008.

[120] P. Poulain, F. Calvo, R. Antione, M. Broyer, and Ph. Dugourd, "Performances of Wang-Landau algorithms for continuous systems," *Physical Review E*, vol. 73, p. 056704, 2006.

[121] D. K. Klimov and D. Thirumalai, "Criterion that determines the foldability of proteins," *Physical Review Letters*, vol. 76, pp. 4070–4073, 1996.

[122] D. K. Klimov and D. Thirumalai, "Linking rates of folding in lattice models of proteins with underlying thermodynamics characteristics," *Journal of Chemical Physics*, vol. 109, pp. 4119–4125, 1998.

[123] J. Kim and Thomas Keyes, "The Influence of Energetic Frustration on the Global Shapes of Energy Landscapes in β-Barrel Forming Model Proteins: Inherent Structure Analysis and Statistical Temperature Molecular Dynamics Simulation," *Journal of Physical Chemistry B*, vol. 112, no. 3, pp. 954–966, 2008.

[124] Robert B. Best, Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal, Michael Feig, and Alexander D. MacKerell Jr., "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ1 and χ2 Dihedral Angles," *J. Chem. Theory Comput.*, vol. 8, pp. 3257–3273, 2012.

[125] A. Barducci, Massimiliano Bonomi, and Michele Parrinello, "Metadynamics," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 1, no. 5, pp. 826–843, 2011.

[126] Christoph Junghans, Danny Perez, and Thomas Vogel, "Molecular Dynamics in the Multicanonical Ensemble: Equivalence of Wang-Landau Sampling, Statistical Temperature Molecular Dynamics and Metadynamics," *Journal of Chemical Theory and Computation*, vol. 10, no. 5, pp. 1843–1847, 2014.

[127] P. Chakrabarti and D. Pal, "Main-chain conformational features at different conformations of the side-chains in proteins," *Protein Engineering*, vol. 11, no. 8, pp. 631–647, 1998.

[128] Gene M. Amdahl, "Computer architecture and amdahl's law," *Computer*, vol. 46, no. 12, pp. 38–46, 2013.

CURRICULUM VITAE

Shanadeen Crystal Begay

Address

Metcalf Science Center
Deparment of Chemistry
Keyes Group, Rm 51
590 Commonwealth Avenue
Boston University
Boston, MA 02215-2521

Cell Phone: (617) 771-3304
Dept Phone: (617) 358-4370
Fax: (617) 353-6466
scbegay@bu.edu, shanadeen@gmail.com

Research Interests

- Extending theoretical biophysics by modeling biologically relevant molecular interactions.
- Incorporating interdisciplinary approaches in computer science, chemistry, physics, and biology to create new simulation techniques for small systems.

Education

Ph.D. in Chemistry with *Certificate in Computational Science,* Boston University, 2015
B.S. Chemistry, Northern Arizona University (NAU), 2006, *ACS-certified*
B.S. Computer Science, NAU, 05/2006, *ABET-certified* both with minor in Mathematics

Scientific Software Skills

- Proficiency in Fortran, Perl, and C/++
- Understanding of molecular dynamics package DLPOLY
- Intermediate experience using *NIX, Windows, various GNU utilities, shell scripting
- Moderate programming with MPI, JAVA, Scheme
- Some experience with HTML, XML, and LATEX, low-level operating systems
- Exposure to the quantum chemistry package GAUSSIAN, LAPACK routines

Research Experience

Graduate Research Assistant, Boston University, Boston, MA 2007 – Present
Advisor: Professor Thomas Keyes
Statistical Temperature Molecular Dynamics simulations with all-atom STMD Methionine Enkephalin and ongoing implementation of the Replica Exchange within the CHARMM code.

Visiting Student, Indiana University - Bloomington, IN 2010 Spring, 1 mo.
Advisor: Professor Peter Ortoleva
Investigated the extension of long time dynamics analysis with extension of the basic theory for order parameter studies in the Virex studies.
The ideas of accelerated dynamics and thermodynamic entropy definitions were developed for langevin dynamics integrators.

Research Assistant, University of Cambridge Chemistry Department, Cambridge, UK 2007 Summer
Advisor: Professor David Wales
Implemented course grain protein potentials into GMIN and OPTIM codes.

Graduate Research Assistant, Boston University, Boston, MA 2005 – 2007
Advisor: Prof. Feng Wang
Began implementation of minimization routine for monte carlo or molecular dynamics study of methane clatherate systems at finite temperature.
ACES Fellowship Recipient.

Research Assistant, Boston University, Boston, MA 2004 Summer
Advisor: Dr. David Coker
Developed and tested methods for treating electronically non-adiabatic relaxation dynamics to determine potential energies for ICN-Zeolite radical reaction using Fortran. Mason Fellowship Recipient.

Publications and Proceedings

- G. L. W. Hart, B. M. Klein, Shanadeen Begay, *Electronic Orgins of Non-stoichiometry and Vacancy Ordering in TiC and TiN,* in Complex Inorganic Solids: Structural, Stability, and Magnetic Properties of Alloys, P. E. A. Turchi, A. Gonis, K. Rajan, and A. Meike (Eds.), Springer (Berlin 2005).
- T. L. Porter, M. P. Eastman, E. Bain and S. C. Begay, *Analysis of Peptides Synthesized in the Presence of Montmorillonite and Cu(II)-Exchanged Hectorite.* Biophysical Chemistry, 91, 115-224 (2001).

Teaching Experience

- Teaching Assistant, Research Writing, Boston University, Fall/Spring 2010/2011
- Teaching Assistant, Discussion sections (3), Chemistry 101, Boston University, Spring 2007
- Mentor, Undergraduate Research Students, BU, Fall 2006

Select Presentations

- *The Thermodynamics and Structure of Methionine Enkephalin using the Statistical Temperature Molecular Dynamics-CHARMM algorithm,* CCP2014 XXVI IUPAP Conference on Computational Physics August 11-14; Boston, USA
- *Statistical Temperature Molecular Dynamics Chemistry at HARvard Molecular Mechanics (STMD-CHARMM) Conformational Analysis of Methionine-Enkephalin (Met-Enk),* Massachusetts Institute of Technology, Jianshu Cao Group, (1/11/13).
- *Navajo Life Ways and Modern Lifestyles: Interpreting the Indigenous Ways into Theoretical Chemical Studies,* University of Massachusetts Boston (U-Mass), Women's Studies Dept., Thurs 12/6/12.
- *Native Americans in Science Yesterday and Today,* Women's Studies course on Female/Minority involvement in science, U-Mass Boston, Amy Den Ouden, Women's Studies 270, Thurs 12/6/12.
- *Conformational Analysis of Small Molecule Systems with Applications to Protein Binding Studies using Statistical Temperature Molecular Dynamics CHARMM* presented at the Greater Boston Area Statistical Mechanics Meeting (09/2012).
- *Conformational Analysis of Small Molecule Systems with Applications to Protein Binding Studies using Statistical Temperature Molecular Dynamics CHARMM* presented at the European Science Foundation Leopold-Franzens-Universitat Innsbruck Conference (07/2012).
- *Statistical Temperature Molecular Dynamics - Chemistry at HARvard Molecular Mechanics (STMD-CHARMM) Conformational Analysis of Methionine-Enkephalin (Met-Enk)* presented at the Northeastern Section Younger Chemist Committee (04/2012).
- *Native Americans in Science Yesterday and Today* presented at the Double Bind Minority Women in Science and Update Thirty Five Years Later Symposium at the American Chemical Society Meeting (03/2012).
- *Gaining an understanding of structure vs. function for Methionine Enkephalin using Statistical Temperature Molecular Dynamics simulations: relationships to folding problems in unstructured domains* presented at Northern Arizona University Chemistry Seminar (11/2012).

- *Deciding on the Undergraduate Commitment* presented at Northern Arizona University Undergraduate Seminar (11/2012).
- *Talking Circle* presented at the 2011 Younger Chemists Committee Leadership Development Workshop (12/2011).
- *Luncheon speaker* at Southwest Regional Meeting Local Section of ACS on Diversity Partner topics.
- *How to succeed in graduate school: Advice to aspiring Juniors and Seniors* Second Annual New England Undergraduate Computing Symposium: Celebrating Excellence and Diversity in Computing
- *Replica Exchange Statistical Temperature Molecular Dynamics* presented at the Tenth Annual Greater Boston Area Statistical Mechanics Meeting (October 18, 2008).
- *Walking the Road Between Two Worlds: Articulating Indigenous Knowledge and Indigenous Rights in Academia* presented at the ACS National Meeting held in Boston, MA (August 20, 2007) and University of Massachusetts Boston (September, 27, 2007).
- *Course Grain DNA potential and Punting!* by Shanadeen Begay, presented at the ACES Seminar held in Boston, MA (September 21, 2007).
- *The Chemical Sciences,* American Indian Science and Engineering Society National Conference held at COBO convention center in Detroit, MI (November 3, 2006).
- *Invesigations of water on the Navajo Reservation,* Minority Access to Research Careers Symposium held in Flagstaff, AZ (April 24, 2004).
- *Uranium and Arsenic Analysis of Water Samples on the Navajo Reservation,* with Jani C. Ingram, presented at ANAS conference held in Phoenix, AZ (April 2003).
- *Pickle Science!* with M. Berry, C. Dominguez, A. Bhuloe, K. Flint held at Flagstaff High School, Flagstaff, AZ (February 6, 2003).
- *Vacancy Ordering in TiC,* Annual Arizona Space Grant Consortium Symposium held at the University of Arizona, Tucson, AZ (April 20, 2002).

Highlights of Career

2013

Native Education Science Initiative (NESI) Advisory Committee Member at Massachusetts Institute of Technology

Organized panels and programming for the Committee of Minority Affairs for National American Chemical Society

Organized the Northeastern section of the Younger Chemists' Committee (NESACS) events

2012

Have completed half of the ACS Leadership training and helped programming on a national level.

Has served on a Boston University Natural Sciences and Curriculum Committee and mentor for students in various leadership positions in many organizations locally and abroad.

Visited Northern Arizona University and Crownpoint Institute of Technology as a visiting scholar to lecture, establish research connections, further service and outreach goals, and begin grant writing collaborations.

Organized the Northeastern section of the Younger Chemists' Committee (NESACS) Career Symposium held 04/2012.

Organized workshop titled Preparing Students for Success in Graduate School by Shanadeen Begay, Boston University and Joslynn Lee, Northeastern University, AISES 34th National Conference, Adaptability.

2009 to current

Mentored for EL Alliance and the ACS Undergraduate Programming
Has retained consistent email communication with these students over the course of their undergraduate careers utilizing best practices in mentoring and teaching pedagogy.

Sequoia Fellow of the American Indian Science and Engineering Society
Organized career development and poster sessions, and seminars, and workshops for members.

2007 to 2012

American Chemical Society National Younger Chemists Committee Member. Organized career development and networking seminars and workshops for members
Coordinated interdepartmental and public communications, facilitated leadership training, member of the Diversity and Inclusion Advisory Board, Career Chair for the Local Boston and VP of BU group(s).
Served as a liaison to many committees within National ACS, assessed leadership skills for pipeline into governance positions with Zenger–Folkman Management training.

2007 to current

Boston University Native American Association (Founder)
Led student team in developing an undergraduate cultural group that serves Boston University

Service Organizations

- American Indian Science and Engineering Society, Massachusetts Institute of Technology, Graduate Student Advisor, Region 6 coordinator, 2006 - current
- American Chemical Society–BU Young Chemists Committee Social Chair, registered Student Member, NAU local member, 2003–2007
- Society of Women Engineers–NAU Chapter President, Treasurer, member of various committees, 2000–2003
- Ambassadors for the College of Engineering and Technology, NAU, 2001–2003
- Accelerated Strategic Computing Initiative Member, 2001–2002
- Minority Access to Research Careers Member, 2003–2005

Fellowships and Scholarships Awarded

- Ruth L. Kirschstein National Research Service Award, 10/2009 - 9/2012
- NSF Integrated Graduate Education and Research Traineeship, 2005–2007
- American Indian Ford Scholarship, 2002–2005
- American Chemical Society Scholarship, 2003–2005
- American Chemical Society Scholar Travel Award, 2006 and 2004
- Lisa Bagwell Engineering Scholarship, 2004–2005

- Raytheon Scholarship, 2004–2005
- Mason Fellowship, 2004
- Proctor and Gamble Scholarship, 2004
- Summer Undergraduate Research Fellowship Award at BU, 2004
- Bull HN Informations Systems Scholarship, 2003–2004
- College of Engineering Industry Committee Scholarship, 2003–2004
- Delia Martinez Memorial Engineering Scholarship, 2003–2004
- National Science Foundation Scholarship, 2003–2004
- Minority Access to Research Careers, 2003–3004
- Ambassadors for the College of Engineering and Natural Sciences Scholarship, 2001–2003
- NASA-NAU Space Grant, 2001–2002
- Chief Manuelito Scholarship, 1999–2002
- Accelerated Strategic Computing Initiative Award, 2000–2001
- NAU Academic Award, 2000
- Sucessful Transition and Academic Readiness Program, 1999–2000

Honors

- Gold Axe Award, NAU, 2005
- Description: Gold Axe Award recognizes seniors who have made outstanding contributions to the university in the areas of academic performance, service, and activities.
- President's Prize, NAU, 2005
- Description: President's Prize is considered the highest student recognition from the university.
- E-mentoring Program and Engineering Day Recognition, EEOP-NAU, 2004
- Undergraduate Research Opportunities Program at BU Award, 2004
- Dean's List for Fall 1999, Spring 2000, Fall 2004
- College Horizons Graduate Program Participant, 2004
- Multicultural Outreach Award, SWE, 2003

- Description: Awarded for organizing Space Camp visit for Navajo mid-school girl from Leupp Middle School and designed circuit building project.
- NAU-NASA Space Grant Intern, 2001-2002
- Accelerated Strategic Computing Initiative Participant, 2000-2001
- Navajo Freshman Year Experience Participant, 1999-2000